# HIERARCHICAL LOOP STRUCTURES REGULATE CHROMOSOME ORGANIZATION

Yunyan He

A dissertation submitted to faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mathematics in the Graduate School.

Chapel Hill
2020

Approved by:
M. Gregory Forest
Kerry Bloom
David Adalsteinsson
Jingfang Huang
Katherine Newhall

## ABSTRACT

Yunyan He: Hierarchical Loop Structures Regulate Chromosome Organization
(Under the direction of M. Gregory Forest, David Adalsteinsson and Kerry Bloom)

The study of chromatin dynamics and motion is essential to the understanding of the rules of life; indeed, the dynamic organization of chromatin plays a unique role in nearly all DNA metabolic processes. It has been hypothesized, and heavily explored experimentally, that various classes of proteins participate in managing the geometrical structure and dynamics of chromosomes throughout the cell cycle. While histones effectively compact chromosomes, the extent of compaction (7-fold) does not account for the greater than 1000-fold compaction required to reduce the genome (length ~1 meter) to the size of the nucleus (radius ~2 micron). It has recently been shown that condensins, one class of structural maintenance of chromosome (SMC) proteins, are able to extrude loops along the chromatin fiber, providing a mechanism to further compact the genome. The role of condensins during mitosis is a potential means for the 10,000-fold compaction attained at the time of chromosome segregation to daughter cells. Condensin and the loop-extrusion mechanism may also be responsible for the heterogeneous shape of chromosomes. In this work, a real-time numerical simulation model is introduced to simulate chromosome dynamics in the presence of histones and condensins to provide a realistic model based on physical principles of polymer dynamics and loop formation. We show that chromosome compaction and variance in compaction result from the coupled hierarchical loop structures created by both histones and condensins, and not from either individual effect.

To study the existence of sub-nuclear domains in the nucleus, another coarse-grained polymer model that incorporates chromosomal crosslinks was examined. The model

replicates the phase separation of the nucleolus, reveals dynamic self-organized clustering of genes within the nucleolus, and accurately predicts that multiple rDNA loci positioned on more than one chromosome will behave remarkably similarly to a single, intact nucleolus, despite varying the geometric positions of the loci. In this work, a new numerical method to identify the 3D territory is developed to measure the volume of the nucleolus. Simulation results reveal that the nucleolus gains compaction due to high frequency binding-unbinding kinetics of dynamic crosslinks, and the multiple-loci nucleolus gains comparable volume with the single-locus case.

## ACKNOWLEDGEMENTS

I would like to thank all of you who helped me or inspired me along the journey of my years at UNC. The impact on me is valuable and will last for my entire life. Forgive me for limiting the list to a few.

*Dad and mom*

Thank you for always supporting me. This has been important to me because you give me the courage and confidence to lead my life in my own way. I feel really sorry for being far apart from you for all these years.

*Yi*

Strange to call you Yi so I will go with Laopo. Thank you for everything, as always. I can carry on because I know you are by my side. The highlight of the past five years was definitely the moment you said “yes”.

*All my friends at UNC*

It has been great fun to develop hobbies with all of you. Basketball, volleyball, video games, chatting … Let’s continue enjoying our lives in the future, no matter where we are.

*My Committee, Jingfang Huang, and Katherine Newhall*

Thank you for guiding my research. A lot of your suggestions are valuable and helped me reexamine my work. I am grateful you agreed to be on my committee.

*David*

The courses and research projects you helped me with showed me the essence of programming and actually led my way towards this path. Thank you very much for your help.

*Kerry and Josh*

I could never image to work with biologists in interdisciplinary study is such a fun and cool thing. Thank you for showing me a different field of science, and thank you for your patience and kind help to a biology-free kid to accomplish all these.

*Greg*

You showed me the strategy to be a great scientist: be vigilant to opportunities and have big scope. This has inspired me and will always lead me in the future. Thank you so much for your guidance.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

3C – Chromosome Conformation Capture

bp – Base Pair

GFP – Green Fluorescent Protein

kbp – Kilo Base Pairs

lacO – Lactose Operon

$L_p$ – Persistence Length

MSD – Mean Squared Displacement

P – Poise

RFP – Red Fluorescent Protein

$R_g$ – Radius of Gyration

SMC – Structural Maintenance of Chromosomes

tetO – Tetracycline Operon

WT – Wild Type

Δ – Null Mutation

# CHAPTER 1: CENTROMERIC CHROMOSOME DYNAMICS SIMULATOR BASED ON POLYMER MODEL


## Summary

The geometrical structure of chromosomes has profound significance in cell biology. Chromosomes are carriers of DNA and are confined and compacted within the nucleus. The mechanistic causes for the compactness and orderliness of chromosomes remains to be explored. In order to learn the physical principles leading to the compact structure, we construct a mathematical simulation pipeline based on a polymer-physics based model to simulate the dynamics of chromosomes inside the nucleus of a living cell. The simulation is numerically stable, and is well-designed for future modifications. The numerical results are tested against existing results in the field of polymer physics. The primary purpose for developing this model is to provide the foundation for constructing genuine simulations involving additional molecular species in the nucleus and provide insights on how chromosomes are condensed.


## Introduction

DNA carries and replicates biological information and is generally accepted as one of the most important genetic materials in living beings. In eukaryotic cells, DNA forms long, compact structures called chromosomes. These structures are highly organized and compacted in the nucleus. There are 23 pairs of chromosomes in each human cell, with a total arclength of roughly 2 meters. These structures are confined in the nucleus by its membrane, with a diameter of about 2 $\mu m$. It requires significant and active compaction processes to keep all chromosomes in this highly condensed state. Hence not only has the DNA

sequencing of nucleotides been a hot topic in cell biology over the past decade, but also the geometrical and topological structure of chromosomes.

In order to study the higher-order structure of chromosomes, biologists seek methods to observe the geometry of chromosomes. The most widely utilized methods to observe chromosomes involve the use of DNA-binding or protein-binding fluorescent markers (Robinett et al., 1996; Straight et al., 1996). Other methods include chromosome conformation capture (3C) (Sati & Cavalli, 2017) and super-resolution microscopy (Neice, 2010). However, there are limitations for these methodologies. Clear observation down to the base pair (bp) scale demands very high resolution, which is not attainable with light microscopes. Moreover, since chromosomes are crowded and tangled in the cell nuclear environment, the difficulty of separating chromosomes and conducting subtle observations on isolated segments arises.

Under this context, it is essential to come up with a numerical simulation pipeline that accurately captures the dynamics of chromosomes in cell environment. Upon a proper choice of model, the simulation mimics the dynamics of chromosomes in arbitrarily high resolution. The time cost is significantly lower compared to the cost of cultivating fluorescently tagged chromosomes for experimental observation. A direct comparison against biological observations can be conducted to benchmark the validity of the numerical model. More importantly, numerical simulations are much simpler and cheaper to test hypotheses, and make predictions to verify experimentally. This creates the opportunity to gain biological insights from simulations about various mutants that change either the property of a chromosome or the environmental factors.

In this work, a specific type of chromosome is studied and simulated: the dicentric plasmid chromosome. It preserves the following properties: it is a small circular DNA molecule; it is physically separated; and it replicates independently. In the Department of

Biology at University of North Carolina at Chapel Hill, Prof. Kerry Bloom's laboratory is able to synthesize plasmid chromosomes, embed tetR-GFP bound to a tandem tetO operator DNA array in the chromosomes at prescribed regions for visualization, and activate the chromosomes in the nucleus in budding yeast cells. The signals captured in microscopy are processed through image processing pipelines to obtain statistics for further analyses. In simulations, the same labeling scheme and image processing pipeline is applied to maintain the best consistency with experiments: the simulation scale is chosen to match the synthesized 10 kbp (thousand bp) plasmid chromosome. It obtains circular geometry with two sites pinned in space to mimic sister centromeres connecting chromosomes to microtubules during the early stage of mitosis; the fluorescent region of DNA is also highlighted in simulation for a direct comparison of simulated microscopy signals with proper transformation to experimentally obtained microscopy signals.

The model is based on the well-defined properties of long-chain polymers in a viscous environment (Rubinstein & Colby, 2003). We establish our model in reference to existing models (Josh Lawrimore et al., 2019) and engage proper modifications. An individual dicentric plasmid chromosome is discretized and simulated as a polymer chain governed by force laws. In addition, fluid drag is incorporated to account for the fluid environment in cells, while not accounting for hydrodynamic interactions, called the free-draining approximation. The choices of force components are subject to polymer physics to guarantee rationality. To validate our simulation, various statistics from the simulation output are analyzed and compared to existing results in polymer physics.

**Simulation Methods**

The simulator was developed in C++ using *DataTank* as the user interface, which allows users to input environmental parameters and visualize the 3D geometrical structure of

simulated chromosomes in this pellucid graphical interface. The C++ program parses 3D coordinates of the chromosome into a time series as the output. The output file can be directly read through *DataTank*, and *DataTank* provides built-in tools for visualization and analyses. An overview of the workflow of the simulation pipeline is provided in Appendix 1. A list of tunable parameters and the default settings are in Table 1.1.

*Model Description*

The simulation is based on a discretized bead-spring model in a fluid drag environment. The total 10 kbp plasmid chromosome is discretized into 386 beads, with each bead representing roughly 10 nm (33 bp) of DNA. Bead 1 and Bead 139 are pinned in space in the simulation to simulate tethered, static centromeres in the nucleus. Beads are connected by springs governed by a linear force law and a bending rigidity law. The persistence length of DNA is 50 nm, set by an additional hinge force in the simulation that controls rigidity. Thermodynamic fluctuations influence the motion of all particles, and the range the substrate explores can be tested as a metric for the intensity of the dynamics and the stiffness of the polymer chain. For faster simulation, a water environment is adopted in the simulation. The fluid environment attains a viscosity of 0.01 P (Poise). Given the estimation of nuclear viscosity to be 141 P, the simulation speeds up the evolution process 14100 times, where 0.15 sec total simulation time is equivalent to approximately 35 min real time.

*Tensile Stiffness*

Two adjacent beads are connected by a spring governed by a linear spring law. It resists extension of the polymer chain between two beads. We simulate the extension by the tensile stiffness of a beam, and thus the equations of stress-versus-strain are used. We do not consider 3D tensors for simplicity.

$$\boldsymbol{\sigma} = E\boldsymbol{\epsilon}, \tag{1.1}$$

$$\boldsymbol{\sigma} = \frac{\boldsymbol{F}}{A}, \tag{1.2}$$

$$\boldsymbol{\epsilon} = \frac{\boldsymbol{\Delta L}}{L_0}. \tag{1.3}$$

The spring force for extension due to tension between two adjacent beads can be calculated by substituting and rearranging equations (1.1) ~ (1.3).

$$\boldsymbol{F_{Spring}} = \frac{EA}{L_0}\boldsymbol{\Delta L}, \tag{1.4}$$

where $E$ refers to the Young's modulus; $A = \pi r^2$ corresponds to the the area of the cross-section of the spring; $L_0$ is the resting length of the spring. For this model, $L_0 = 10\ nm$ since each bead represents about 10 nm of DNA; $E = 2\ GPa$ and $r = 0.6\ nm$, consistent with the plasmid chromosome we use.

*Bending Rigidity*

Chromosomes have a bending rigidity that is associated with DNA linkers. DNA linkers are the naked DNA strands linking adjacent nucleosomes (a nucleosome is a substructure formed by ~150 bp of DNA wrapping around an octet of histone proteins). One standard metric to describe the rigidity in polymer physics is the persistence length, the length over which correlations between motion of two individual ends are lost (Rubinstein & Colby, 2003). Greater persistence length refers to the fact that the motions at two distant locations are still closely correlated, and thus indicating stiffness over the respective distance between those locations. The rigidity of our chosen plasmid chromosome is approximated by that of B-form DNA, which is reported to be 50 nm (Dekker et al., 2001). The entropic flexibility of DNA is consistent with the worm-like chain model in polymer physics (J F Marko & Siggia, 1995; John F. Marko & Siggia, 1994) which makes worm-like chain models a common choice in simulating the dynamics of DNA. However, for our simulation focusing on discretized DNA segments smaller than the scale of the persistence length, the DNA

bending rigidity deviates from worm-like chain models. Hence, the worm-like chain model is not applied in our model.

The bending rigidity for the polymer chain is embedded by applying additional force at each of the bead, in order to keep the substrate straight (Bloom, 2008). If a deflection occurs between two connecting springs, a shear bending force is produced at the joint to eliminate the deflection. The mechanism is achieved by considering the chromatin as an elastic beam, and then Euler-Bernoulli beam theory is applied.

To simplify the system, instead of solving the Euler-Bernoulli equation, a coarse-grained model is chosen. Assume an elastic rod of length $2d$ attains a load at the center, and the rod is perfectly stiff on both halves from the middle. The restoring force is then proportional to the degree of deformation characterized by the angle two halves of the rod make, $\theta$ (Young et al., 1976).

$$\boldsymbol{F_{Hinge}} = \frac{2EI}{d^2}\theta\boldsymbol{n}, \tag{1.5}$$

where $E$ and $\boldsymbol{n}$ refers to Young's modulus and the normal vector along the direction between the corresponding springs, respectively. $I = \pi r^4/4$ is the moment of the area of the beam with circular cross section. $d$ is chosen to be $L_0$ in this simulation to achieve $L_p$ of 50 nm for the polymer. The name "hinge" comes from the hinge-like mechanism the rigidity force is modeled after. See Figure 1.1 for graphical illustration.

$\boldsymbol{F_{Hinge}}$ is applied at every bead where the two connecting springs fail to bend straight, in the direction pointing inward from the angle the springs form. $-\boldsymbol{F_{Hinge}}/2$ is applied at each of the beads adjacent to the center bead, in the opposite direction, to prevent overall displacement caused by bending rigidity.

In simulation, the angle $\theta$ can be calculated by the geometric information of the spring. To be more precise, $\theta = \pi - \arccos\left(\frac{\boldsymbol{v_1} \cdot \boldsymbol{v_2}}{\|\boldsymbol{v_1}\|\|\boldsymbol{v_2}\|}\right)$ where $\boldsymbol{v_1}, \boldsymbol{v_2}$ are the vectors from the center bead to two adjacent beads, respectively. A fast approximation algorithm for the inverse cosine function is adopted for fast simulation.

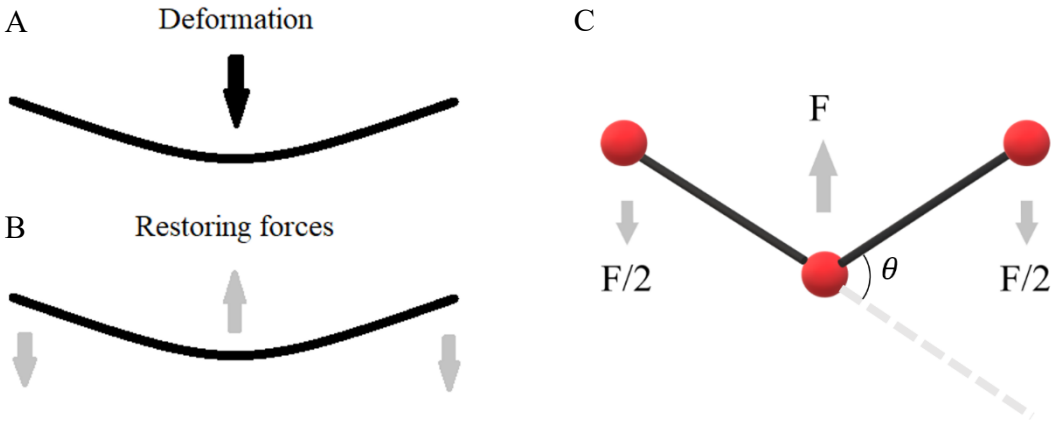



*Figure 1.1: Illustration of hinge force*
(A) Intuition of hinge force inspired by Euler-Bernoulli beam theory. The chromosome segment is modeled as an elastic beam under deformation at the center. (B) The restoring forces resist the deformation. (C) The discretized form of restoring force used in the simulation, translated from continuum model.

*Thermal Fluctuations*

The random motion of small particles under thermodynamics leads to stochastic behavior for the system. In the simulation the Brownian-like motion is modeled as the consequence of random forces applied on each of the bead and then averaged out for a small time period, $\Delta t$. For a freely diffusing particle in 1D, the root mean square displacement is proportional to the square root of $\Delta t$.

$$\Delta x_{RMS} = \sqrt{2D\Delta t}, \tag{1.6}$$

where *D* is the diffusion constant determined by the fluid environment. Einstein relation explains that the diffusion is proportional to the energy the system possesses and the mobility of the fluid

$$D = \mu k_B T, \tag{1.7}$$

where $k_B$ refers to the Boltzmann constant, $T$ is the temperature for the simulation, and $\mu$ is the mobility of the fluid, generally known as the ratio of the particle's terminal drift velocity to the corresponding force applied,

$$\mu = \frac{v_d}{F} = \frac{dx}{Fdt}. \tag{1.8}$$

For a cell environment simulation, because the fluid obtains low Reynolds number down to around $Re_{cell} \approx 6$, laminar conditions are triggered and the mobility is essentially the inverse of the drag coefficient:

$$\mu = \frac{1}{\zeta}. \tag{1.9}$$

Moreover, Stokes law is applied for diffusion of spherical particles in a laminar fluid condition, which gives

$$\zeta = 6\pi\eta r, \tag{1.10}$$

where $\eta$ is the viscosity of the medium, and *r* represents the radius of the spherical particle. Therefore, rearranging equations (1.7) ~ (1.10) leads to the Stokes-Einstein relation to simulate the random fluctuation

$$D = \frac{k_B T}{6\pi\eta r}. \tag{1.11}$$

If we substitute equation (1.11) into the mobility equation (1.6), the standard deviation of the random force that simulates thermal fluctuations is obtained:

$$SD = \sqrt{\frac{12\pi\eta R k_B T}{\Delta t}}. \tag{1.12}$$

Here $\eta$ is a viscous environmental parameter. Since we simulate the model in a water environment to accelerate the dynamics, water viscosity is chosen. R is chosen to be the diameter of a single bead (equivalent to bead separation). We simulate the thermal force by randomly generating a component in each dimension drawn from a normal distribution with mean centered at zero and variance to be the square of $SD$, $F_{thermo,i} \sim N\left(0, \frac{12\pi\eta R k_B T}{\Delta t}\right), i = 1, 2, 3.$

*Collision Repelling Force*

It is intuitive that two arbitrary segments of the chromosome are not capable of crossing and overlapping. In simulations, the same mechanism is under consideration. Two methods are implemented: the bead-collision model and the cylinder-collision model.

In the bead-collision model only bead-bead collisions are considered. To prevent two beads that are distant along the polymer chain from colliding with each other, an additional force is applied for every pair of beads geometrically nearby. This collision force is governed by a linear Hookean spring force between beads when they intrude in the territory of the other beads. The repulsive force becomes

$$\boldsymbol{F_{collision}} = -k_c(2R_c - \|\boldsymbol{\Delta x}\|)\boldsymbol{n}, \text{ if } \|\boldsymbol{\Delta x}\| < 2R_c, \tag{1.13}$$

where $k_c$ and $R_c$ are the collision constant and collision radius, respectively. $\boldsymbol{n}$ represents the normal vector of $\boldsymbol{\Delta x}$ from the target bead to its neighbor. $R_c$ is chosen comparable to the radius of beads. Hence the collision force is not triggered when two beads are not overlapping. This model does not strictly prevent situations of chromosomes passing through one another.

In the cylinder collision model, the major difference is that springs are also not allowed to invade another spring's volume. In simulations, each spring obtains a virtual cylinder bounding box, and the collision is triggered when cylinder contact occurs. The repelling force is generated when two cylinders collide, and the principle for the cylinder repelling force is the exact linear Hookean force as is modeled in the bead-collision model. The force generated from the body of the cylinders is then distributed to both end beads of the cylinder by conserving the moment of the force.

$$\boldsymbol{F} = -k_c(2R_c - \|\boldsymbol{\Delta x}\|)\boldsymbol{n}, \text{ if } \|\boldsymbol{\Delta x}\| < 2R_c, \tag{1.14}$$

where $\boldsymbol{n}$ represents the normal vector to the plane formed by two colliding springs. $\boldsymbol{\Delta x}$ is the distance of the springs in the 3D space. Then the force is distributed to both ends by preserving the moment to prevent the cylinder from spinning:

$$\begin{aligned} \boldsymbol{F_1} &= \frac{L_2}{L}\boldsymbol{F} \\ \boldsymbol{F_2} &= \frac{L_1}{L}\boldsymbol{F}. \end{aligned} \tag{1.15}$$

Here $L$ represents the total length of the target spring, while $L_1$ and $L_2$ are the lengths from the collision site to one of the two ends of the spring, respectively.

In practice, the bead-collision model is adopted to simulate the collision force. Three dominating reasons lead to the decision. Firstly, the bead-collision model requires significantly less amount of computation since no further computation is needed to be spent on detecting cylinder collisions. Secondly, throughout testing process the cylinder-collision

model did not show great statistical influence on the performance of the model. Most importantly, due to the function and abundance of particular enzymes such as topoisomerase II, chromosomes can "cut" the chromatin chain and cross one another. This rationalizes the concept of crossing chromatin chains.

*Fluid Drag Force*

The drag force applied on each bead when moving in the fluid environment is again governed by Stokes law. The force scales with the velocity and is applied in the negative direction of the velocity to prevent the particle from moving.

$$\boldsymbol{F_{drag}} = -\zeta \boldsymbol{v}, \tag{1.16}$$

where $\zeta = 6\pi\eta a$ refers to the drag coefficient of spherical particles in fluid with viscosity $\eta$ in a small Reynolds number system. The radius of particle *a* is comparable to the diameter of the beads. $\boldsymbol{v}$ is the velocity of the particle.

*Equation of Motion*

In an inertial system, the equation of motion for each individual particle subject to all forces listed above is:

$$\begin{gathered} m^{(i)} \frac{d\boldsymbol{v}^{(i)}}{dt} = \boldsymbol{F}^{(i)}_{\boldsymbol{spring}} + \boldsymbol{F}^{(i)}_{\boldsymbol{hinge}} + \boldsymbol{F}^{(i)}_{\boldsymbol{thermal}} + \boldsymbol{F}^{(i)}_{\boldsymbol{collision}} + \boldsymbol{F}^{(i)}_{\boldsymbol{drag}} \\ \frac{d\boldsymbol{x}^{(i)}}{dt} = \boldsymbol{v}^{(i)} \end{gathered} \tag{1.17}$$

where $\boldsymbol{x}^{(i)}$, $\boldsymbol{v}^{(i)}$ and $\boldsymbol{F}^{(i)}$ are the position, velocity and corresponding force for the *i*-th bead in the simulation, respectively. One can solve the system of stochastic differential equations by iteratively updating velocity and position.

However, the equation of motion can be simplified considering the system to be drag-dominated system instead of an inertial system. In low Reynolds number environment,

moving objects experience a drag force shown in (1.16), and will reach a terminal velocity. The equation of velocity in the fluid writes

$$\boldsymbol{v}(\boldsymbol{t}) = \boldsymbol{v_t}(1 - e^{-\frac{t}{\tau}}) = \frac{\boldsymbol{F}}{\zeta}(1 - e^{-\frac{t}{\tau}}), \tag{1.18}$$

for an arbitrary force $\boldsymbol{F}$. Here $\tau = \frac{m}{6\pi\eta r}$ is a small constant time. $e^{-\frac{t}{\tau}}$ decays to 0 rapidly as time proceeds. In this limit, any motion of objects caused by external force $\boldsymbol{F}$ can be approximated as a uniform motion at the terminal velocity $\boldsymbol{v_t} = \frac{\boldsymbol{F}}{\zeta}$. Then the equation of motion can be approximated by a straight-forward stochastic ordinary differential equation

$$\zeta \frac{d\boldsymbol{x}^{(i)}}{dt} = \boldsymbol{F}^{(i)}_{\boldsymbol{spring}} + \boldsymbol{F}^{(i)}_{\boldsymbol{hinge}} + \boldsymbol{F}^{(i)}_{\boldsymbol{thermal}} + \boldsymbol{F}^{(i)}_{\boldsymbol{collision}} \tag{1.19}$$

This equation of motion gets rid of acceleration, which greatly enhances the stability of the numerical system. Moreover, both memory and time is saved because of the reduction of calculation for every iteration.

*Stability and Constraints*

The model utilizes an explicit Newton solver with dynamic time step to ensure both stability and efficiency. The motion among all particles for each time step is strictly bounded in order to prevent divergence.

Two beads labeled by 0 and 138 are modeled as centromeres and are assigned fixed coordinates in space. Other than this, there are no additional geometrical confinement conditions applied. The nucleus membrane is not considered in the simulation, due to the fact that the chromosome is pinned in space by two spindle pole bodies roughly 800 nm apart in the pericentric region of the nucleus, which is far away from the nucleus membrane. The plasmid chromosome is so small compared with the 2 $\mu m$ nuclear diameter on average that the necessity to take interactions between nuclear membrane and the chromosome into

consideration is minimal. A table of tunable parameters and their default values are listed in Table 1.1.

*Simulation Program*

The simulation is coded in C++, using *DataTank* as its graphical interface. All input parameters need to be initialized in the *DataTank* script. After initialization, *DataTank* passes all parameters and configuration file to C++ compiler, and starts the simulation. The program keeps track of the 3D positions for all particles. In each iteration, the simulation program calculates forces included in the system at each time step, gathers and sums all forces for each particle, calculates the displacement for the particle over the time step, and evolves the positions of the beads accordingly.

The program parses the 3D geometry of the polymer chain and saves the positional data into the output file for a prescribed time period. The output data can either be read and displayed by *DataTank* directly, or is also allowed to be exported to be analyzed by other numerical analyses tools. Real-time observation is also viable in *DataTank* if the simulation is running locally. For detailed manual, see Appendix 1.

*Table 1.1: Plasmid chromosome simulation tunable parameters*

| **Parameter** | **Default value** |
|---|---|
| Bead resting separation $L_0$ | 10 nm |
| Temperature $T$ | 25℃ |
| Viscosity $\eta$ | 1 centipoise |
| Young's modulus $E$ | 2 GPa |
| DNA sectional radius $r$ | 0.6 nm |
| Collision force damping factor $k_c$ | 0.25 |
| Collision radius factor | 0.9 (of bead separation) |
| Drag dambing radius factor | 0.8 (of bead separation) |
| Time step | 2 ns |

**Results**

The current model does not include necessary components yet for simulating chromosome dynamics. It is the intermediate model in the process of constructing full simulations that can be directly compared to biological observations. However, there are existing results the model can compare to in order to validate the program. There are well-established theories in polymer physics on the dynamics of a polymer chain under entropy fluctuations, which is known as the Rouse model (Rubinstein & Colby, 2003). Comparisons can be conducted to the Rouse model on which this model is based, with only few modifications. A sample simulation output is shown in Figure 1.2.

With thermal fluctuations the only external source of force, the bead-spring system diffuses as a random coil after it reaches its equilibrium state. This process is robust since initial geometry does not influence the equilibrium. However, strong regulation can be visualized in the beginning stage, compared to fluctuating particles. Particles located at the center of the chain are subject to strong restrictive forces by the neighboring beads, and thus performs strong sub-diffusive dynamics. On the contrary, the behavior of particles at both ends of the chain more incline to pure Brownian Motion.

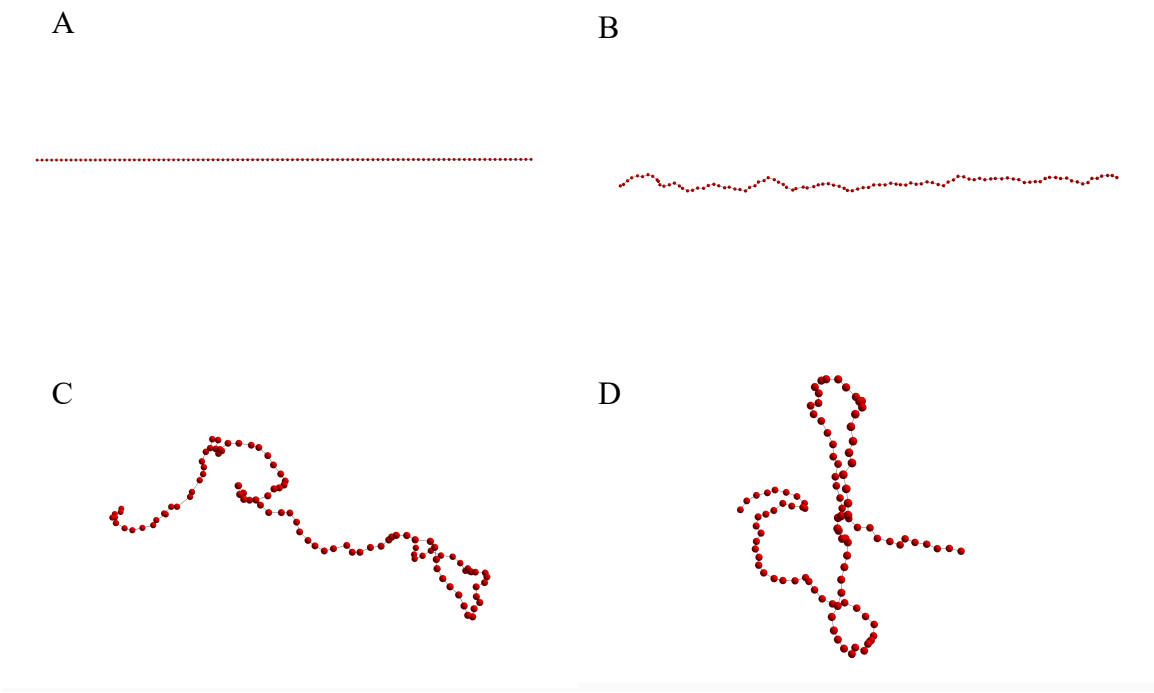



*Figure 1.2: Simulated free-floating linear polymer chain*
Illustration of simulated free-floating linear chain. Each red bead corresponds to a discretized chromosomal polymer unit. The connecting springs are shown as black line segments connecting the beads. The example simulation contains 100 beads. (A) Initial geometry $t = 0\ s$. 100 beads are located 10 nm away from the nearest neighbors at the resting length. (B) $t = 3e^{-5}\ s$. (C) $t = 3e^{-3}\ s$. (D) $t = 0.03\ s$

*Brownian motion validation*

The first step for validating the model is to test if the method of introducing thermal force to represent Brownian motion for each bead is valid. In the simulation, a random force is applied to each particle to simulate an identical dynamical displacement Brownian motion leads to. This can be simply tested by generating multiple unconnected, free-floating particles in the simulation. The collision force is turned off, making thermal force the only source of external force for the dynamics. The mean squared displacement (MSD) from the simulations are compared to the expected values of a particle experiencing Brownian motion (Rubinstein & Colby, 2003):

$$\langle r^2 \rangle = \frac{k_B T}{\pi \eta R} t \tag{1.20}$$

where $\eta$ is the viscosity of the fluid and $T$ is the temperature. *R* refers to the radius of the bead. $k_B$ is the Boltzmann constant. Hence the MSD scales linearly with the elapsed time *t*.

This well-established result is used to confirm the validity of our thermal force term in the model. Statistical tests have been applied to guarantee that the numerical design is consistent with analytical results with probability above 95%.

*Rouse time*

One established analysis for Rouse models is the mode decomposition of the polymer. In the Rouse model, the diffusion of a polymer chain in thermal dynamics is represented by Brownian motion of beads connected by harmonic springs. Researches have shown that, despite the initial geometry, the Rouse Chain will collapse into a random coil after a particular amount of time from any initial configuration. Moreover, the time required to reach the steady state can be calculated, given the environmental parameters, the polymer stiffness and the size of the polymer (Rubinstein & Colby, 2003).The lowest order Rouse mode relaxation time, which is called the Rouse time, is calculated by

$$\tau_R = \frac{\eta b^3}{\pi k_B T} N^2 \tag{1.21}$$

Here $N$ is the number of segments and $b$ is the Kuhn length of the polymer. The parameters corresponding to the environment are the following: $\eta$ is the viscosity of the fluid; $T$ is the temperature. $k_B$ is the Boltzmann constant.

*End-to-end distance*

Another famous analysis for Rouse models is the end-to-end distance of the polymer, which measures the length of the end-to-end vector of the polymer. After the polymer reaches its equilibrium state and fluctuates as a random coil, the end-to-end distance obeys a certain distribution. The mean of the end-to-end distance only depends on the stiffness and the size of the polymer (Rubinstein & Colby, 2003). The mean end-to-end distance is

$$\langle R \rangle = \sqrt{Nb^2} \tag{1.22}$$

for a Rouse model consisting of *N* particles, where *b* is the Kuhn length of the polymer (Kuhn length = $2L_p$).

*Radius of gyration*

One of the other commonly studied statistics in polymer analysis is the radius of gyration which is defined as

$$R_g^2 = \frac{1}{N} \sum_{k=1}^{N} \|\boldsymbol{x_k} - \boldsymbol{x_{mean}}\|^2 , \tag{1.23}$$

where *N* is the total number of particles on the chain. $\boldsymbol{x_{mean}}$ represents the mean position of all the particles, and thus the right-hand side of the equation is the averaged squared distance from all the particles to their mean position. Intuitively, the radius of gyration $R_g$ captures the compactness of the polymer at an instance.

This metric can be plotted over time as a proof of validation. There is close relationship between $R_g$ and end-to-end distance. The expectation of the radius of gyration is

$$\langle R_g \rangle = \sqrt{Nb^2/6} \tag{1.24}$$

for a Rouse model consisting of *N* particles.

The simulated outputs are compared with analytical results by comparing radius of gyration and end-to-end distance. Statistics from simulated results are averaged over samples from 3 independent runs. The expected end-to-end distance and radius of gyration are calculated from equations (1.22) and (1.23). See Figure 1.3, Figure 1.4 and Table 1.2 for comparison results.

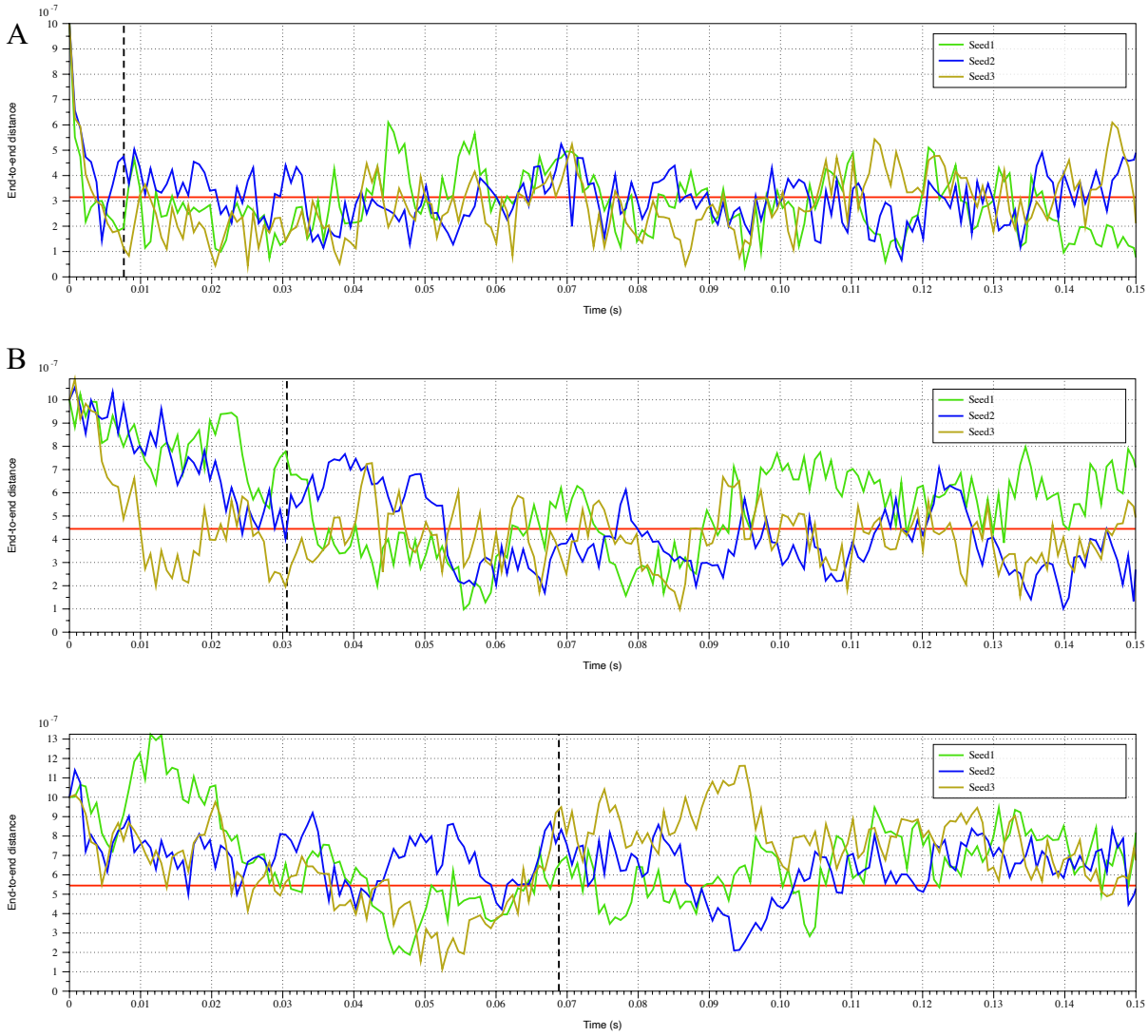



*Figure 1.3: End-to-end distance plot of simulated free-floating linear chains*
End-to-end distance plot for (A) $1\ \mu m$ linear chain, (B) $2\ \mu m$ linear chain, (C) $3\ \mu m$ linear chain. Red horizontal line refers to analytical end-to-end distance of the polymer governed by (1.22). Black dashed vertical line is the analytical Rouse time given by (1.21). All 3 runs for each experiment are independent, and the plot is constructed over 200 sampled instances for each run.

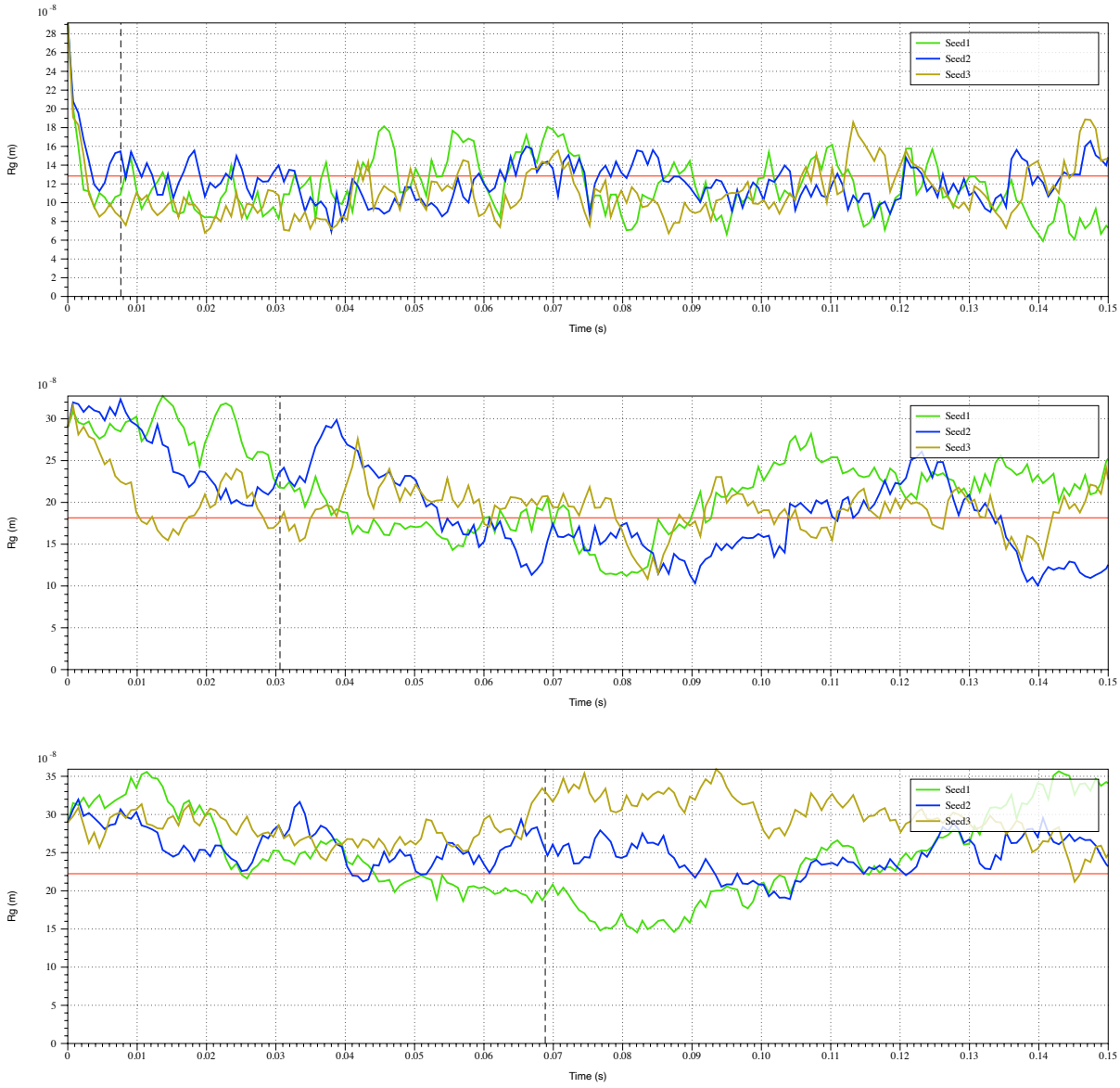



*Figure 1.4: Radius of gyration plot of simulated free-floating linear chains*
Radius of gyration plot for (A) 1 $\mu m$ linear chain, (B) 2 $\mu m$ linear chain, (C) 3 $\mu m$ linear chain. Red horizontal line refers to analytical $R_g$ of the polymer governed by (1.23). Black dashed vertical line is the analytical Rouse time given by (1.21). All 3 runs for each experiment are independent, and the plot is constructed over 200 sampled instances for each run.

*Table 1.2: Averaged radius of gyration and end-to-end distance*

| | **Radius of gyration (nm)** | | **End-to-end distance (nm)** | |
|---|---|---|---|---|
| | Expected | Simulated | Expected | Simulated |
| 1 $\mu m$ chain | 129 | 122.17 | 316 | 284.14 |
| 2 $\mu m$ chain | 183 | 194.63 | 447 | 421.53 |
| 3 $\mu m$ chain | 224 | 244.17 | 548 | 568.87 |

The statistics shown in Table 1.2 are measured and calculated through samples from simulation outputs to validate the simulation model. The simulations all start with linear geometry. The equation governing the system is the same equation shown in (1.19), except for that the collision force term is excluded from the equation, which is consistent with the fact that the Rouse model does not take collision into consideration.

The good agreement with the Rouse time, expectation of end-to-end distance and expectation of radius of gyration calculated in theory confirm accuracy of the numerical code in the Rouse chain limit. Notice that even though the model is based on Rouse chains, there are subtle differences in the force terms this reasonable deviation is acceptable.

**Discussion**

The polymer chain model is the basis for simulations for chromosomes. It stands out in delivering fast real-time simulations of chromosome dynamics, provides alternative approach to visualize the chromosomes in nucleus, and raise hypotheses that enlighten the field of cell biology. To achieve this, additional modules need to be implemented and added to the model in order to imitate real chromosomes. For instance, it is widely accepted that SMC proteins play an important role in organizing chromosome structures, and thus it is necessary to construct models for those proteins and embed them into the model. Moreover, the chromosome obtains different geometrical structures and nuclear environment during various stages of the cell cycle, meanwhile the default geometry simulates the chromosome during the early stage of mitosis. Thus, programs simulating the dynamics of chromosomes in other stages of the cell cycle could be achieved with modifications on the geometrical constraints and environmental parameters.

Researchers have discovered profound knowledge on the fundamental structures of chromosomes. However, simulating the dynamics in the cell environment still requires

imagination. There have been discussions on the choice of models for simulating chromosomes. The choice of both the fundamental model and the influencing components should be closely dependent on the scale of the particular model. For instance, to simulate the dynamics of the entire genome in a budding yeast nucleus at a coarse scale, the worm-like chain is utilized and the hinge force is neglected (Vasquez et al., 2016, Hult et al., 2017). For this work, the simulation focuses on a finer scale of a single plasmid chromosome, down to the finest scale of 10 nm. Therefore, polymer stiffness is relevant and a linear Hookean approximation between beads is reasonable since beads do not get far enough apart to require a nonlinear stiffening condition. There are possibilities for choices of separate fundamental principles, for example a non-linear spring force law, to be discovered and to be tested under resolved experimental conditions.

# CHAPTER 2: HISTONES AND CONDENSINS COMPACT CHROMOSOMAL STRUCTURES


## Summary

The chromatin fiber wraps around an octet of histone proteins to form a nucleosome, which is regarded as the basic structural element of DNA packaging. SMC proteins, including condensin and cohesin, have been studied for their versatile function in compacting and organizing the structure of chromosomes. They are indispensable components during the process of chromosomal compaction. In this work, numerical models for condensin, one major class of SMC proteins, and histones are carried out and embedded to the simulation pipeline. They interact with DNA and create hierarchical loop structures along the chromatin. The models for the proteins are constructed based on published experimental results, and their functionalities are tested. The simulation results draw critical intuition into the mechanism by which condensins and histones synergistically regulate the chromosome structure, keeping the chromosome structure compact meanwhile dynamic.


## Introduction

The geometrical structure of chromosome is of prodigious significance in cell biology. The mechanisms that enables long chromosomes to be packaged into nucleus have attracted a tremendous amount of attention. It has been discovered that the vast amount of DNA was regulated in space not only by its double-helical spatial structure, but also maintained by functioning proteins and polymer fluctuation in nuclear environment.

There are three major classes of proteins that manages the geometrical and topological shape of chromosomes: the DNA topoisomerases, which participates in the linking and

unlinking of intertwined DNA strands; the SMC proteins, which includes various classes of proteins that take part in multiple aspects of DNA organization; Last but not least, the DNA-binding proteins, which binds to DNA at their DNA-binding domains.

Among these proteins, we focus our interest on the particular types of proteins that specifically contribute to DNA compaction. While DNA topoisomerases manage the topological structures of the DNA supercoil in the winding point of view, this property is parting from the scale and concentration of our simulation; In the family of SMC proteins, there are two typical classes of proteins: cohesin and condensin, which have become popular topics of research in cell biology in the past decade. While cohesins involve in sister chromatid cohesion during DNA replication, we have focused on condensins, which implied by their name are an essential factor in chromosome condensation; In various classes of DNA-binding proteins, we fix our attention to histones, the protein that forms the basic structural unit of chromatin in eukaryotes, because histones are also considered as one of the key elements in chromosome packaging and compaction in the cell nucleus.

SMC proteins represent the family of complexes that participate in various forms of chromosome organization. Among this large family of chromosomal ATPases, condensins are large protein complexes that stand out specifically for chromosome condensation, and have become a hot topic among researchers in the field. Not only their biochemical composition, but also the dynamics they bring to the chromatin structure that have been carefully studied. Each condensin complex is composed of five subunits, among them three non-SMC regulatory subunits (one kleisin subunit, Brn1 and two HEAT-repeat subunits, Ycs4 and Ycg1) and a pair of core SMC subunits, known as SMC2 and SMC4. Studies have shown that condensins possess a wide variety of chromosome functions. For instance, condensins contribute to the formation of chromosome territories (Bauer et al., 2012). The mechanistic cause for the functions is studied in a microscopic scale. A more recent study

shows that the condensin behaves as a mechanochemical motor that moves along DNA (Terakawa et al., 2017), and there is biological evidence for loop extrusion hypothesis (Ganji et al., 2018).

Although the existence and biochemical constitution of condensins have been confirmed, how condensins help chromosome in condensation is still vague. Since a huge amount of chromosome as well as an approximately equal concentration of proteins huddle inside the nucleus, it is less possible for direct observation of interaction between condensins and chromosomes in living cells. Alternatively, researchers separated DNA from nuclear environment and successfully visualized the dynamics of DNA chain with condensin activated in vitro. By applying a constant flow to the DNA for stretching, researchers clearly observed a DNA loop structure formed by a condensin, and moreover the loop size grew gradually at a varying rate. This behavior is called the loop extrusion of condensins. Although this experiment showed condensins' activity in full detail, the fact that the experiment was not conducted in cell environment weakens the power of the experiment. The condensin dynamics in the nucleus still remains uncertain.

Unlike condensins, histones form stationary subunits by acting as spools around which DNA winds. This leads to shortened length (~7 times fold) of chromosomes. These proteins are composed of 2 nearly symmetrical halves, essentially 4 histone proteins (Luger et al., 1997). Each nucleosome is a highly stable unit. Besides chromosomal compaction (Johansen & Johansen, 2006), histones also participate in chromatin regulation in diverse processes, such as DNA repair (Supek & Lehner, 2017) and gene regulation. However, for tethered chromosomes during early mitosis, both experiment and simulation suggest that histones do not significantly decrease the signal size of the marked region of dicentric plasmid chromosomes. How histones compact chromosomes in particular geometrical constraints needs to be explored.

Condensins and histones are claimed to be proteins that contributes to chromosome packaging and condensation. Yet the detailed demonstration that explains the underlying mechanistic basis remains veiled. Here we take another approach to learn how condensins and histones transform the properties of the chromosome polymer and condenses the strand. We construct models for histones and condensins intuited by experimental results and involve them in the polymer chain model. Tests and comparisons with statistics from experiments need to be conducted to validate the model. Then we will take a glimpse at the structural changes the proteins deliver to the system, and clearly observe the mechanism of how histones and condensins compact and condense the chromosome structure.

## Methods

### *Histones*

Experimental observations have provided clear observation report in vivo on the structure of histones at a high resolution (Luger et al., 1997). On average 147 bp of DNA wraps 1.65 times around the histone octamer at a very high occurrence along the chain. Essentially histones attach to the chromatin spine every 200 bp across all eukaryotic genomes. Then these histones build higher order structures by forming assemblies among nucleosomes and stabilized by linker histones. Study have also shown that histones obtain intrinsic periodical turnover. Extensive histone exchange happens in nucleus, independent of cell replication. Furthermore, histone protein H3 behaves as a tension sensor that helps regulating the cell during mitosis (Luo et al., 2016). DNA unwind from the histone protein spool under high tension during the dividing process. These are both mechanisms for histones to self-regulate. Moreover, histone synthesis is reported to be dynamically regulated by the modulation of the half-life of histone mRNA (Morris et al., 1991), which implies that histone enrichment is heterogeneous in nuclear environment.

Histones in our simulation are modeled according to published results (Luger et al., 1997). In the simulation, histones are modeled as multiple stationary 7-bead subloops along the chain (each unit represents $\sim 30\ bp\ \times\ 7\ =\ \sim 210\ bp$ size of nucleosome). Those subloops are formed by applying virtual springs at the attachment sites. The polymer is fully loaded by histones by default in simulation due to high enrichment of nucleosomes in observations. Each histone unit periodically attaches to and detaches from the chromatin chain to simulate the turnover of histones observed in vivo. The turnover mechanism is achieved by setting an intrinsic timer to each unit. One histone unit switches status once the timer exceeds the predefined deadline. In simulation, in order to stochastically vary the average number of active histones, one can tune either the total number of histones in the simulation or the on/off period of the histones. Moreover, winding DNA under high tension is capable of unwinding histones, which is consistent with observations of force-dependent unwinding and rewinding mechanism.

In the simulation neither the mass of histones nor the higher order structural interactions between nucleosomes are considered upon this point. We focus instead on the consequences of the loop structures. Figure 2.1 A shows an example of the geometry of the chromosome while interacting with a histone. The ring-like substructure remains static until the histone detaches from the DNA.

*Condensins*

The biochemical composition of condensins have been well studied (Hirano et al., 1997). They are known as the key component in cell organization and compaction. During metaphase, the spindle axis is enriched with condensins (Stephens et al., 2011). A natural assumption on the behavior of condensins is that condensins help to organize the DNA geometrical structure, in order for the genome to successfully replicate and become evenly

distributed into the daughter cells during mitosis. However, the underlying mechanistic function condensins perform when interacting with the chromatin remains unknown, with a number of hypothesis and corresponding models illustrating different possible explanations (Sakai et al., 2018; Thadani et al., 2012).

In a previous study, condensins were modeled as static springs (J Lawrimore et al., 2016). These protein complexes are able to form large subloops along the chromatin chain. Later, with Terekawa et al. published their examination of the behavior of single condensing complexes on DNA sheets, condensins are confirmed to obtain mobility while they are actively creating loops (Terakawa et al., 2017). In their work, condensins are regarded as biological motors that translocate along the DNA at a rate of about 60 bp/s.

A recent work by Ganji et al showed the motion pattern of condensins, which they name it the loop extrusion (Ganji et al., 2018). By utilizing real-time imaging techniques, they are able to visualize the behavior of a single condensin while binding to a separated, flow-stretched DNA in vitro. Provided energy, the condensin complex is capable of creating a loop on that DNA and continuously expanding that loop. Additionally, they found that the extrusion behavior does not possess a constant rate. The extrusion slows down as the DNA is further stretched, which is explained as another influential hypothesis: the loop extrusion rate of condensins depends on the tension along the binding DNA. This publication has been revolutionary in the field, leading to numerous downstream studies.

We developed a prototype condensin model that is able to translocate along the DNA chain while extruding loops. Intuitively, each condensin complex in the model can be considered as a person grabbing a rope. As the person is grabbing the rope towards him, an expanding loop will form. When the person experiences an extreme tension along the rope, a sudden release is triggered. To be more precise, the condensins are modeled as a dynamic spring that can connect two beads in the polymer chain. The two ends of condensins are

heterogeneous. One end is more dynamic in the sense that it constantly reaches out to distant segments along the chain. We call this end “the weak end”. Once the weak end lands on the chromosome, the core spring pulls the attachment site towards the other end, which is called “the strong end”, and thus a loop is formed. The strong end is relatively stable. It generally remains stationary, only releases the chain once the condensin core spring attains too much tension. After detachment, the condensin shrinks to its initial configuration and attach to the chain at a segment closer to the weak end. This mechanism allows the condensin complexes to form a loop and translocate simultaneously. This model is similar to established condensin model (Josh Lawrimore et al., 2017). However, the performance of this model is not optimal. It fails to reconstruct the extremely compact chromosome structure.

An up-to-date model of condensin complexes we construct is based on the loop extrusion observation published by Ganji et al. The main difference from our previous model can be interpreted also by the “rope pulling” analogy. In this model, the person does not release the rope upon excessive tension. The hypothetical model of such a complex is a ring structure that anchors on a segment of DNA. The rest portion of DNA is grabbed and pulled by the complex through its ring structure and completes the extrusion activity. In our model, each complex is modeled as a spring attaching to the DNA strand, with one end anchored at a fixed site, which is called ‘the strong site’. The strong end does not release upon high local tension, comparing to our previous model. The other attachment site of the complex, ‘the weak site’, moves along the chain, attach to a farther segment of the DNA, and then the spring pulls the two ends close again to complete the extrusion activity. A customized function governs the rate at which the weak site takes one action in order to simulate the decaying extrusion rate shown in experiments. Moreover, after a predefined period of extrusion, the spring releases and restart to extrude another loop. Instead of a multi-step relaxation process, the loop releases and disappears instantly. This mechanism is in

accordance with the sudden disruption of the condensin loop in experiments. Figure 2.1(B) - (D) illustrate how a condensin loop evolves in a short period. The extrusion rate was amplified compared with real simulations, for a clear observation of the loop extrusion behavior.

The histones and condensins we introduce into the simulation do not conflict with each other. Each bead can be winding around histones while being a portion of the condensin extruded loop at the same time. The proteins are able to attach to every bead, except the two end beads of the circular chromosome which are used to model sister centromeres. The program will parse the coordinates for all beads and save them into the output file, same as the output from the model in Chapter 1. However, one can choose to output the positions of the attachment sites of the proteins at every saving stride for further studies. By default, the indices of condensins' attachment sites are also written into output file. A list of tunable parameters on histones and condensins are listed in Table 2.1.

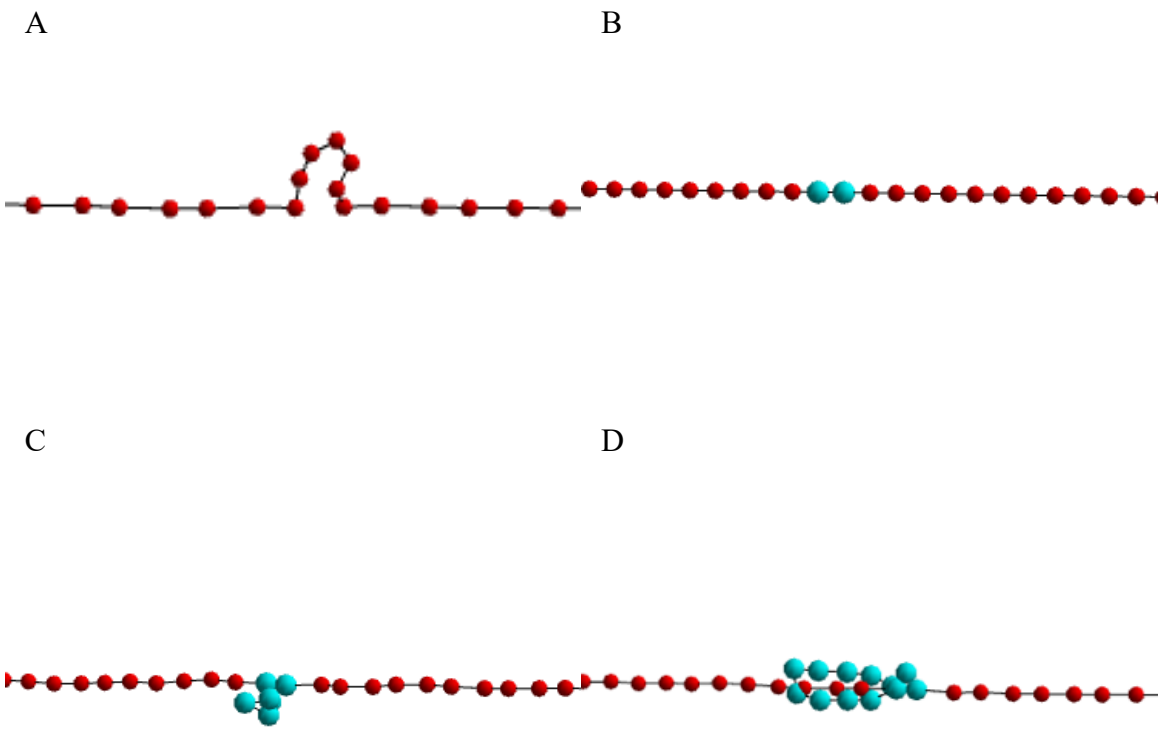



*Figure 2.1: Illustration of histone model and condensin model*
(A) In simulation, a histone creates a small, static loop at its attachment site. (B) Initialization of condensin model. The blue beads represent the substructure formed by a single condensin complex. (C) The condensin starts to extrude a loop. (D) The condensin forms a large, dynamic loop.

*Table 2.1: Tunable parameters corresponding to DNA-binding proteins*

| **Parameter** | **Default value** | **Parameter corresponding class** |
|---|---|---|
| Number of histones | 50 | Histone |
| Histone loop size | 7 | Histone |
| Mean On | 5 $ms$ | Histone |
| Mean Off | 0 $s$ | Histone |
| SD | 0.1 $ms$ | Histone |
| Detach | Off | Histone |
| Detach threshold | 30 $nm$ | Hisone |
| Number of condensins | 6 | Condensin |
| Dynamic extrusion | On | Condensin |
| Judge time | 20 $\mu s$ | Condensin |
| Increment | 10 $nm$ | Condensin |
| Extrusion decay rate | 1 | Condensin |
| Detach | 1 | Condensin |
| Mean On | 11.3 $\mu s$ | Condensin |
| SD On | 4.3 $\mu s$ | Condensin |
| Mean Off | 0 $s$ | Condensin |

*Data analysis methods*

In the Bloom Lab at Department of Biology, University of North Carolina at Chapel Hill, timelapses of a fluorescent-tagging synthesized dicentric plasmid chromosome was obtained and used for analyses. The chromosome is isolated and circular with two centromeres when it loses the ability to replicate (autonomously replicating sequence, ARS) and becomes an ideal target for chromosomal analyses. It is labeled with a 5.5 kbp tetO/TetR-GFP array (Dewar et al., 2004) for observation purpose (Figure 2.2 A). The chromosome is circular, with two centromeres located asymmetrically, creating a longer arm and a shorter arm. These two portions of the chromosome are called “the loose arm” and “the tight arm”, respectively. The GFP portion is located in the middle of the loose arm.

Original biological data are a series of signal images from the fluorescent portion of the dicentric plasmid chromosome. The image series are obtained from parsed time lapses every 30 seconds for 20 minutes in vivo. Due to the restriction of resolution, the images are blurry bright spots. Characteristics of the chromosome can be represented by certain statistics from the images. Typical statistics include measuring the “signal length”, which is generated by a custom MATLAB GUI (J Lawrimore et al., 2016). In short, the signal length is the major axis length of the minimum oval covering the signal. In some cases, the “signal width”, the minor axis length of the oval, is also taken into consideration.

Moreover, we observe frequent oscillation in the size of the signal in vivo. To examine the signal variations, we create the density map to show change in signal length as a function of temporary signal length. The change in signal length is measured by the difference between the current signal length and the signal length observed one following timestep. The frequency is displayed by color code where red is the most dense and blue is the least dense.

In order to directly compare simulation data with biological observations, we convolve the corresponding GFP-tagging fluorescent portion of simulation result with a point-spread function using the program called Microscope Simulator 2 (Quammen et al., 2008). This method allows to convert simulation results which are exact positions to vague signals with great consistency. Then the simulated images are analyzed using a custom MATLAB program that automatically converts the simulated images into binary masks and uses the REGIONPROPS function in MATLAB to measure the signal lengths. The resulting statistics can be directly compared to the experimental results.

## Results

### *Experimental Results*

The signal length histogram (Figure 2.2) reveals a right-skewed distribution, with mean and median between 600-700 nm. The most frequently observed signal length is between 500-600 nm, showing a very compact structure of the signal. The density heatmap in addition shows frequent signal fluctuation by several observations located off the center of the y-axis. The fit for the distribution supportively suggests the skewness of the distribution (Figure 2.2 C). Instead of a symmetric Gaussian distribution, a bimodal 2-term Gaussian distribution fits the data better.

The plasmid signals reveal the characteristics of the dynamics of plasmid chromosomes: condensed while dynamic. The compactness is essential in chromosomal organization through cell cycles. Meanwhile, the mobility of chromosome is indispensable. Studies have shown the close correspondence between chromatin motion to nuclear fields, such as DNA damage (Zimmer, 2018) and genome stability (Dion & Gasser, 2013).

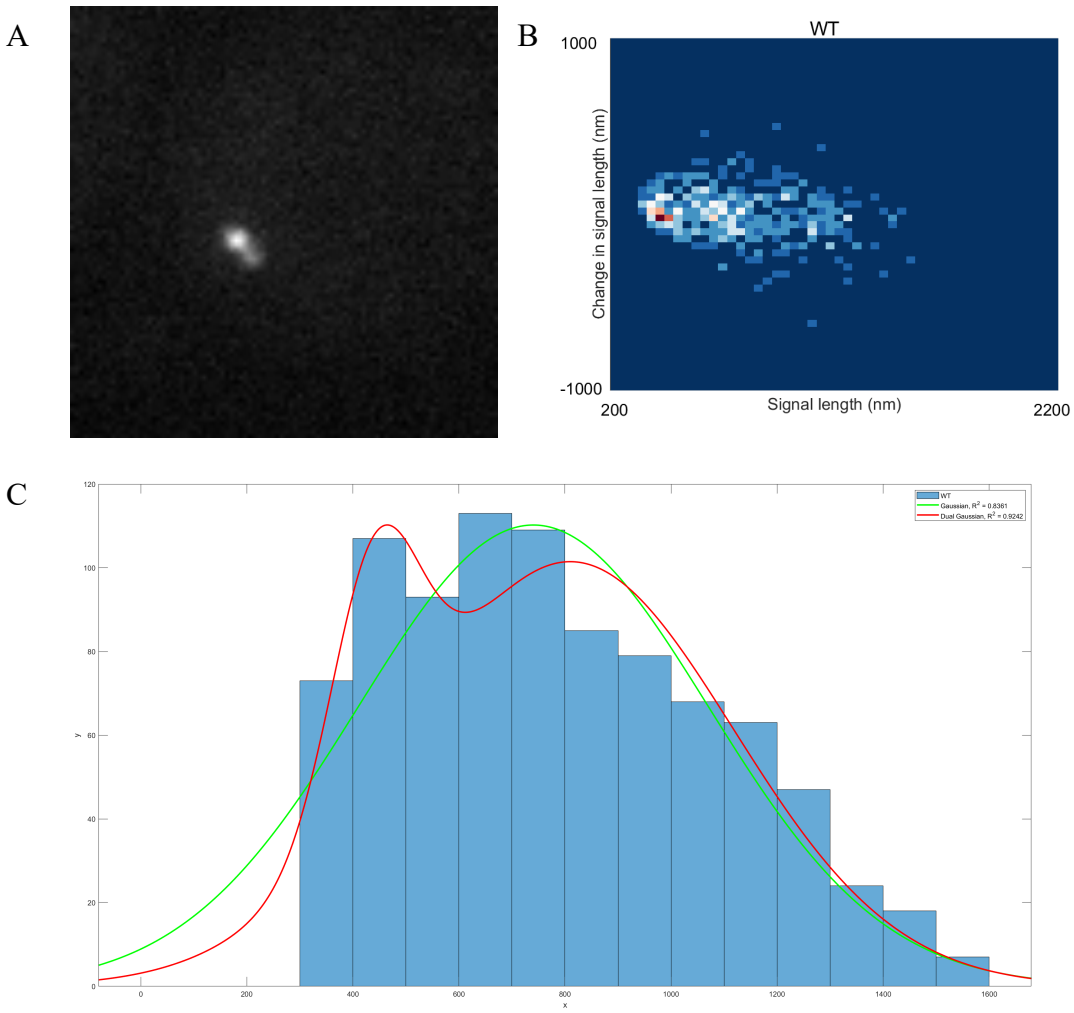



*Figure 2.2: Dicentric plasmid chromosome signal and signal length analyses*
(A) The signal of WT plasmid chromosome with tetO/TetR-GFP array. (B) Density heat map of change in signal length plotted against signal length. (C) Histogram of plasmid chromosome signal length in normalized frequency. The histogram is also fitted with one-term and two term Gaussian distributions. The fit is evaluated through $R^2$ statistics. Greater statistics implies better fit.

*Simulation Results: histones and condensins compact the dicentric plasmid*

To investigate the functions of the structural proteins of interest, we compare the simulation outputs by gradually add components to the system and observe the intermediate results. That is, we compare the signal length from 4 distinct systems: the raw polymer chain system introduced in Chapter 1; the polymer chain system with histones; the polymer chain system with condensins; and finally, the polymer chain system with both histones and condensins. We extract the simulated signal length by similar approach as in experiments, transform the signals to acquire the metrics and directly compare to the statistics from experimental results. Reasonable errors are tolerable between simulated results and experimental results. The ultimate goal is to observe similar properties revealed by experiments: condensed signals with intensive variation.

The raw polymer chain model shows large signal lengths, with rare fluctuation (Figure 2.4 A, B). The simulated signal lengths are symmetrically distributed and are most frequently observed to be between 800 and 900 nm, combining up to $\pm$150 nm of variation. While the thermal fluctuation is acting as the only external source for energy, the system lacks the ability to oscillate at a comparable scale to real chromosomes. Moreover, the large averaged signal length suggests that additional components is influencing the system and compacting the structure.

When histones are introduced, we observe a stable structure with very little fluctuation (Figure 2.4 C, D). The compact geometry created by histones tighten the chain, which lead to less overall fluctuation in space (Figure 2.3 B). In other words, histones stiffen the polymer chain by creating substructures. However, histones alone fail to reduce the signal size. Observations with 800-900 $\pm$ 50 nm signal lengths occupy over 90% of the total observations. Compact signals, such as the small round signals observed in vivo, never appear in simulation, even with histones fully loaded. Clearly the two-site-tethered

geometrical structure of the plasmid leads to this consequence. One can imagine the effective compacting effect of histones for free-floating chromosomes in other cell cycles than mitosis. However, there must be supplementary mechanisms that lead to the condensed structure of dicentric plasmid chromosomes.

When condensins are included in the model, the signals show very different dynamics. The vital distinction is the intense fluctuation. One can observe the signal oscillation through both the geometry and the plots (Figure 2.3 C, Figure 2.4 E, F). In the evolution movie of the polymer, condensins keep extruding loops, creating traversing large scale substructures. These active motors bring energy to this system, causing more observations with extreme structures. Through the histogram, a great amount of data spread out to either small signal length down to 400 nm or large signal length up to 1200 nm. This signal variation confirms with experimental data. However, the averaged signal length is not shifted. Most of the signals are still 800-900 nm in length, and the distribution shows no right-tailed skewness.

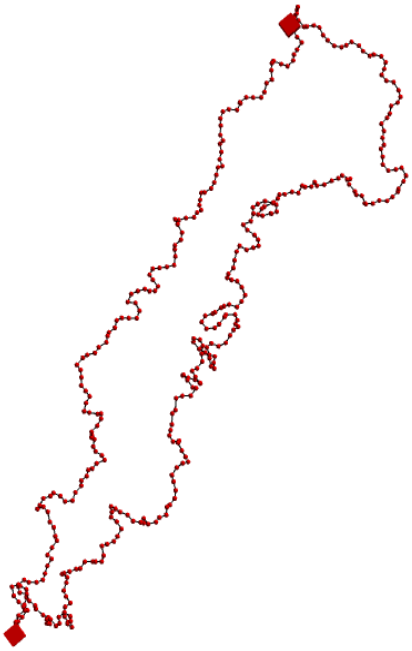
The best approximation to experimental results comes from simulations combining both histones and condensins. Both key factors, the condensed structure and the intense fluctuation, are satisfied. The most frequent signal lengths are around 500 nm, and the distribution is right-skewed (Figure 2.4 G). The heatmap shifts its center from 800 nm to 500 nm while maintaining the vertical spread (Figure 2.4 H). Also, it maintains the ability to create extreme observations up to 1200 nm. The box plot shows direct comparison of averaged signal lengths (Figure 2.4 I). The synergetic mechanism of histones and condensins successfully reduce the signal length while maintaining mobility. Presumably, the mobility is guaranteed by condensins, since introducing condensins alone will greatly amplify the variation of signals.

The geometry shows straight forward intuition on how condensins and histones succeed in condensing the chromosome structure (Figure 2.3). Condensins create large loop structures. Instead of wildly diffusing, the large loops were stiffened and condensed by substructures formed by histones. These large loops will become dense bulbs (Figure 2.3 D). Mean squared displacement (MSD) will drop significantly for these loops when histones are loaded on them, which indicates that, in cases where the fluorescently tagging region is interacting with condensins, the signal will be dense. Meanwhile, the extrusion behavior is able to stretch the chromosome spine. When the fluorescent region is not interacting with any condensins, the signal will be stretched together with the stretched chromosomal spine, which are strong candidates for large signal observations. Consequently, we observe oscillating signals, and the chromosome signals are able to achieve uncommon situations, either remarkably small or extremely large. The overall compactness of the chromosome is obtained also by these large loops, and is fortified by the tension-dependent extrusion mechanism of condensins cooperating with histones. When histones and condensins are both activated and large loops have been formed, high-tension regions will be created along the main spine. These high-tension areas restrict the extruding activity of condensins, forcing them to be relatively stable. Therefore, the system is more inclined to be condensed. This explain why more observations are displayed as compact signals.

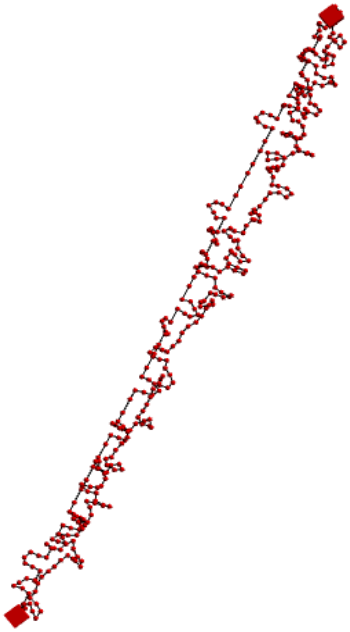
In conclusion, for dicentric plasmid chromosome, the condensed structure is created by the synergetic compacting effect of histones and condensins. Condensins bring fluctuation to the system by actively extruding loops along the chromosome, and also create large dynamic substructures. Histones, on the other hand, function as local buffer to compact those large loops formed by condensins. These two DNA-binding proteins, which are considered as two primary class of proteins that compact and package DNA, lead to a condensed yet

fluctuating chromosomal structure, reconstructing conclusions from microscopy observations.

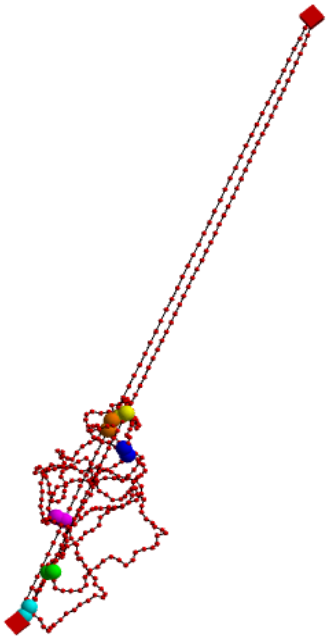
A



B



C



D

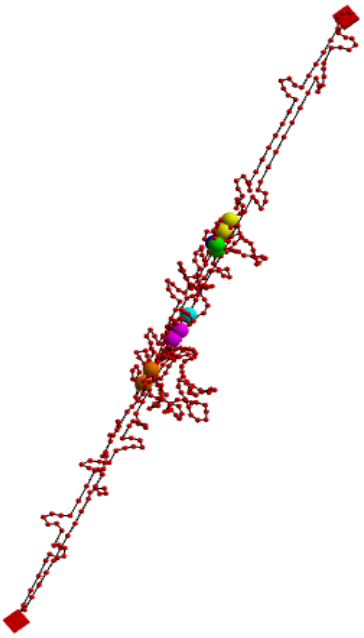
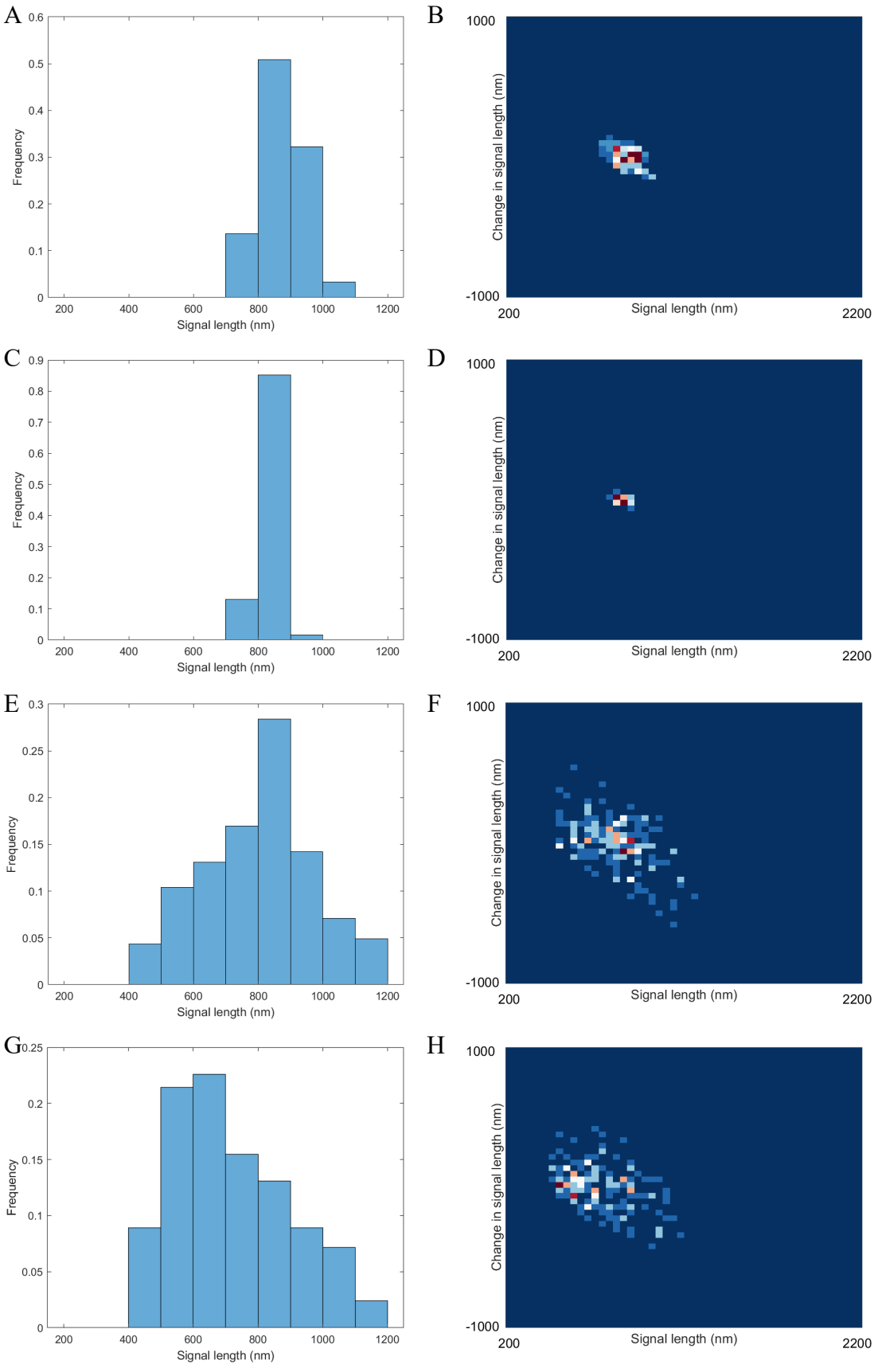


*Figure 2.3: Geometry of simulated plasmid chromosome*
3D Geometry of simulated polymer model at time $t = 0.15s$ under distinct conditions of structural proteins. Small red beads are the discretized unit and are connected through springs, which are black segments in figures. Two red boxes represent centromeres and are fixed in space. (A) Geometry of raw polymer chain model. (B) Geometry of polymer model with histones. Little loops along the strand show locations of histones. (C) Geometry of polymer model with condensins. Large spheres with the same color are the attachment sites of one condensin complex. (D) Geometry of polymer model with both histones and condensins.


A
Frequency
Signal length (nm)
B
Change in signal length (nm)
Signal length (nm)
C
Frequency
Signal length (nm)
D
Change in signal length (nm)
Signal length (nm)
E
Frequency
Signal length (nm)
F
Change in signal length (nm)
Signal length (nm)
G
Frequency
Signal length (nm)
H
Change in signal length (nm)
Signal length (nm)

I

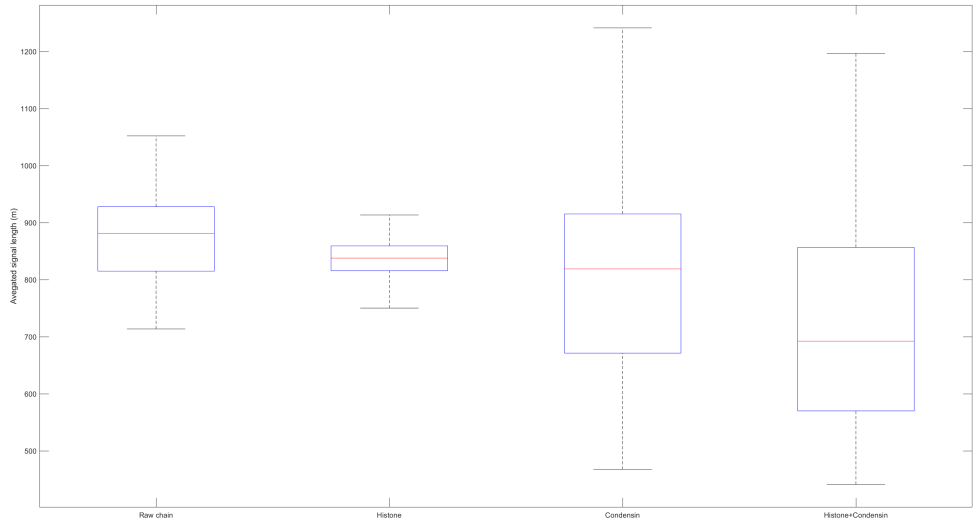



*Figure 2.4: Signal length analyses of simulated plasmid signals*
Comparisons of histograms and density heatmaps of simulated plasmid signal length against types of protein included in simulations. Data for each figure was a total of 183 timelapses selected from 3 independent simulation runs. Timelapses was chosen only after the steady state has been reached. (A) Signal length histogram of raw polymer chain model. (B) Density heatmap of raw polymer chain model. (C) Signal length histogram of polymer model with histones. (D) Density heatmap of polymer model with histones. (E) Signal length histogram of polymer model with condensins. (F) Density heatmap of polymer model with condensins. (G) Signal length histogram of polymer model with both histones and condensins. (H) Density heatmap of polymer model with both histones and condensins. (I) Signal length box plot as a function of distinct classes of proteins included in the simulation. The red lines represent mean, the blue boxes show quantile, and the error bars show SD.

As expected, the geometrical confinement significantly impacts the result. Clues can be found through the analyses of tension along the polymer chain. The local tension can be deduced from spring length, which will be stretched in balance of tension. Figure 2.5 A, B show typical snapshots of tension in different protein conditions. The following conclusions can be drawn: Firstly, the proteins will form low-tension regions within the loops. Meanwhile, since sub-domain structures occupy a large portion of the chromosome, the remaining backbone will obtain high tension propagated from the tethering sites. Secondly, the tension is hardly built even on the backbone without either histones or condensins. Even with condensins extruding large loops, the backbone does not show significant tension without histones. Similar conclusion can be induced from Figure 2.5 C. Although tension has been built by histones on the tight arm (left), the geometrical configuration prevents any local tension on the loose arm (right). The springs stay in their resting length. Since our fluorescent region is on the loose arm, neither histones nor condensins alone are capable of transforming the dynamical properties of the plasmid signals to observational standards.

However, when histones and condensins are both introduced, high tension has been built along the polymer chain, which reveals the strong compacting effect the synergy possesses. The remarkable tension suggests the significant geometrical deformation and modification caused by the proteins under the tethering confinement.

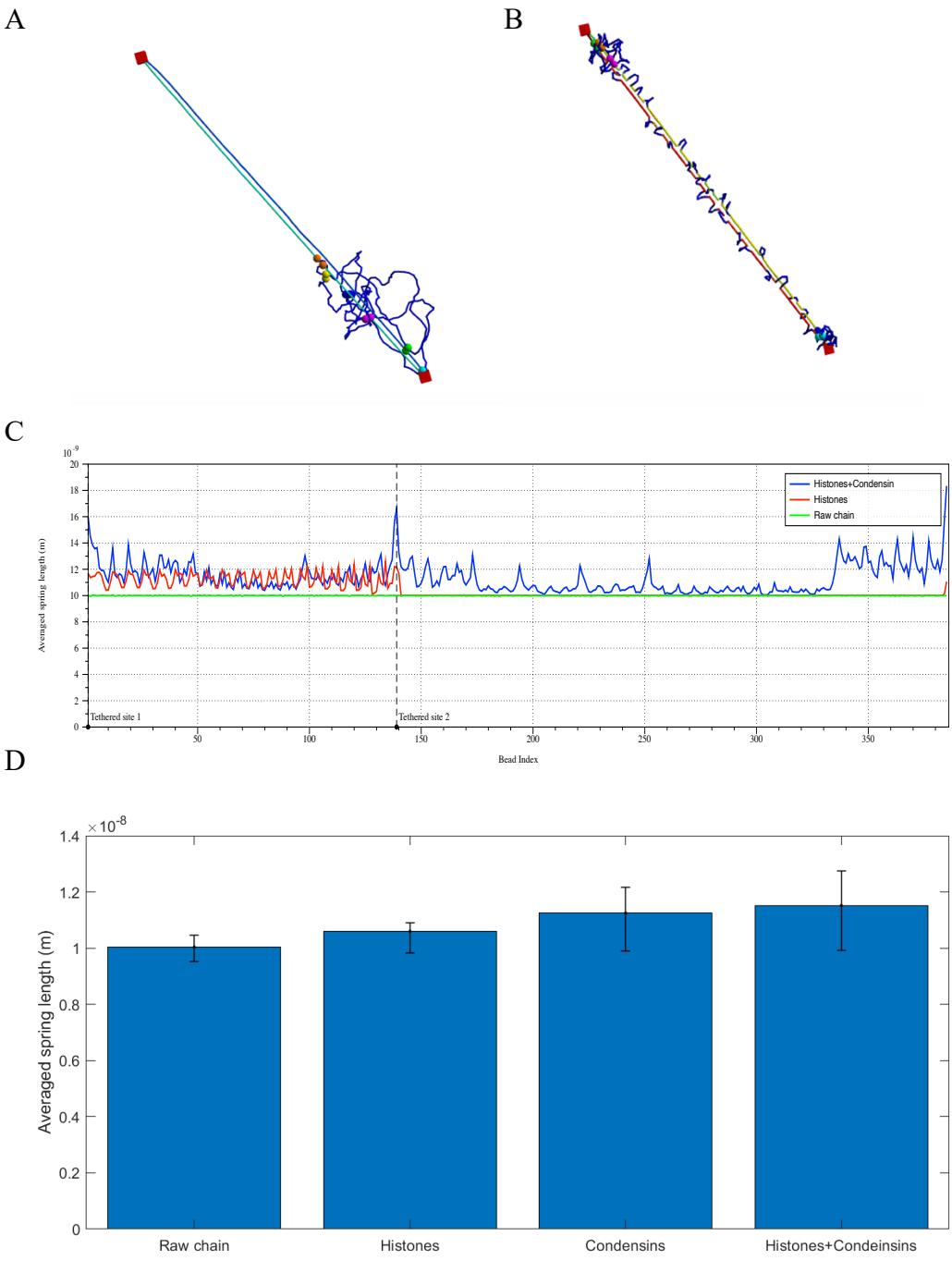



*Figure 2.5: Tension-related analyses of the plasmid chromosome in various protein conditions*

Color-coded images represent tension for plasmid simulations in the following conditions: (A) Including condensins; (B) Including both histones and condensins. Blue is low tension and red is high tension. (C) Averaged spring length plot as a function of bead index. Two centromeres (Bead 0 and Bead 138) are labeled by the dashed lines. (D) Emsemble average of spring lengths as a function of different protein conditions. Error bars show SD.

**Discussion**

After including histones and condensins, this model is capable of reproducing critical results that mimics the dominant statistics from biological observations. Hence the model greatly enhances its validity. However, we cannot claim the validity of the model, since fitness of statistical analyses does not imply mechanistical validation until direct biological observations are able to show the exact dynamics of the chromosome. There are other models that potentially fit biological intuition and worth testing.

One modification is also based on loop extrusion model, yet with a different mechanism. In our condensin model, the condensin complex starts its loop extrusion behavior by reaching its weak end out to a farther particle. The target bead can be arbitrary bead, as long as the geometrical position is the closest to the target position. In real simulations the target bead is not often the next bead of the previous attachment site along the chain. This behavior in simulation result in occasional discontinuity during loop extrusion. For example, the weak site might translocate from bead 5 to bead 50 in one step.

However, we have no direct evidence to explain how exactly does the loop extrusion happen in vivo. An alternative model may forbid the condensin translocating discontinuously. In such an implementation, the condensin is only allowed to translocate to its successive neighboring bead. In the experiment published by Ganji et al., a constant flow is applied to the chromosome to make the chromosomal geometry visible in microscopy. Provided this advection, the chain remains a fine linear structure in space. Both models work in such a condition. However, the environment in nucleus is not alike. The chromosome is naturally twisted and entangled due to its complicated structure as well as the noisy surroundings. Presumably, the two models would perform very differently, and biological observations fail to provide direct evidence to show which implementation is superior. A numerical experiment indicates that our condensin model fails to extrude large loop on low persistence

length substrates, because thermal fluctuation dominates the dynamics. In consequence, the chain fails to maintain its linear structure. In this situation, our condensin model often fails to proceed to successive beads and instead jumping around in a small region. Yet there have been reported observations where condensins releases the large loop spontaneously and start to extrude a new DNA loop at the same location (Ganji et al., 2018), which means that the condensin occasionally releases the extruded loop for some reason, rather than disassociate from the DNA. It has been considered as the primary support for our choice of condensing model.

Another promising modification based on this model is to incorporate fluid dynamics into the system, and better depict the geometrical structures of the structural proteins. We can model histones and condensins as masses, and all the masses interact with each other as well as the fluid surroundings. This definitely will greatly help simulating the real nuclear environment, and will lead to interesting intro- and inter- chromosomal interactions propagates by fluid. Also, condensins can be better simulated as energy-consuming motors by introducing specific momentum to the condensin complex. Histones, furthermore, will obtain its 3D structure that allows DNA to wind around it. By further development following this path, we anticipate an optimal simulation that models the dynamics of the entire nucleus.

Meanwhile, the challenge is obvious. One needs to embed an efficient solver for Navier-Stokes equation into the system that solves the dynamics of fluid with a large number of objects interacting with the fluid. Moreover, the dynamics of histones and condensins also need to be taken into consideration, which significantly expands the scale of computation. The required numerical consumption is tremendous. Although our model neglects some detailed components, it acts as a fair starting point which provides us with a glance of how hierarchical loop structures formed by histones and condensins work to package DNA in living cells.

# CHAPTER 3: HIERARCHICAL CHROMATIN LOOPS REGULATE CHROMOSOMAL STRUCTURE IN VARIOUS WAYS


## Summary

Nuclear structural proteins have been considered one of the essential components that lead to chromatin organization and compaction. Besides packaging and condensing the chromosome, proteins form high-order chromosome organization in adaptation to diverse needs during cell cycles. Previously we simulated two major classes of structural proteins, histones and condensins, as morphological units causing chromatin loops. The experiment-inspired simulations reveal qualitative and quantitative consistency with observational results and establish their value for further studies. In this chapter, we will extensively explore the impact caused by hierarchical chromatin loops. To compare, experimental data of mutated yeast genome offers variation targeting at specific classes of proteins and provides ground truth in vivo. Furthermore, existing studies in polymer physics relate the mobility of the polymer to the stiffness of the material. In this work we study the mobility of the polymers by tuning the stiffness of the chromosome polymer chain while maintaining the hierarchical loop structures. Through this study we are able to show that the stiffness of the chromosome chain is only one factor that controls high-order chromosome organization and dynamics, another being dynamic protein crosslinkers.


## Introduction

The study of chromatin dynamics and motion is crucial, due to the fact that the organization of chromatin plays unique roles in nearly all DNA metabolic processes. A wide variety of methods and techniques have been utilized to examine the motion and structure of

chromosomes. Biologists attempt to observe the motion directly with high-resolution microscopy of live cells. For instance, chromosome motion can be visualized through a single-particle tracking technique using Fluorescent Reporter-Operator Systems (FROS) (Chalfie et al., 1994). Polymer-physics-based simulations provide an alternative approach that has been widely accepted. Regardless of the methodology, the idea that random Brownian motion fails to explain sophisticated chromosomal dynamics has been developed (Bornfleth et al., 1999; Oshidari et al., 2018; Weber et al., 2012). Many additional factors influence the dynamics of chromatin, including the interactions between the chromatin and environmental factors such as proteins, nuclear membrane, neighboring chromosomes and so on.

We have developed a simulation pipeline for the chromosomal dynamics in budding yeast. In previous chapters, a numerical model simulating a dicentric plasmid chromosome based on a polymer bead-spring chain model has been carried out. Furthermore, interactions between chromosomes and particular classes of DNA-binding proteins have been included in the modified simulation to enhance the performance in simulating real chromosomes. Inspired by biological observations, these functional proteins are modeled as either static or dynamic crosslinks that form loop structures on the chromatin. In contrast with raw polymer-chain models, the polymer system generated by this modified simulation shows compactness and mobility, which is more consistent with experimental results in vivo. Hence, the loop structures caused by functional proteins significantly impact the organization and motion of the chromosome. Further investigation of these sub-domain structures is shown valuable in order to elucidate mechanisms influencing chromatin dynamics.

*Hierarchical loop structures cause condensed yet dynamic chromosome organization*

In Bloom's Lab at Department of Biology, UNC-CH, SPT10 Δ experiments have been conducted on chromosome strains to knock out the SPT10 gene in the yeast genome. The yeast SPT10 gene is reported as histone H3 acetylase with a role in transcriptional regulation and sequence-specific activator of histone gene (Natsoulis et al., 1994). The mutation SPT10 Δ will cause a depletion of histone synthesis, leading to significant drop in histone concentration (Eriksson et al., 2005). Unexpectedly, the depletion of histones does not lead to significant decrease in the averaged size of the chromosome, rather it causes more extreme variation in dynamic properties. More signals from the GFP on dicentric plasmid chromosomes are observed to be either extraordinarily large or small. Provided that histones affect the packaging and organization of DNA, this observation infers that under certain geometrical constraints as these plasmids, histones do not condense chromosomes independently.

We are able to reconstruct similar condition of SPT10Δ phenotypes in our simulation by tuning histone-related parameters to decrease histone concentration. Histones are modeled as static regional loops along the polymer chain. They periodically attach to and detach from chromatin, causing local structural compaction. While controlling other conditions, the depletion of histones can be simulated and the cause for unexpected extreme observations can be closely analyzed.

Another experiment designed and conducted by the Bloom Lab is *ycg1-2* mutation. *ycg1* is one of the components of the condensin complex that possesses chromatin-binding function. The mutant *ycg1-2* is temperature-sensitive. When they are released from G1 arrest at restrictive growth temperature (37℃), the chromatin binding effect is weakened and eventually leads to depletion in attached condensin on the chromosome (Li et al., 2011). In contrast with the control group where the mutants are arrested at permissive temperature

(24℃) where the signal distribution is similar to WT simulation, the experiment group shows a clear bimodal signal distribution (Figure 3.4). In observations of dicentric plasmid chromosomes, depleted condensins lead to an increase in averaged signal length. This increase is mainly caused by a new, shifted mode in the frequency histogram. Meanwhile, the frequency mode found in the control group remains detectable. Clearly different from the control group, the observational result displayed by the test group suggests that dysfunction of condensins leads to a complicated mechanistic deviation.

Since resolution restriction prevents researchers from learning the mechanism of condensins directly in vivo, numerical simulations provide a valuable exploratory tool. In our simulation, condensins actively extrude asymmetrically along chromatin, causing large dynamic loops. The dysfunction of condensin in *ycg1-2* can be simulated by modifying parameters that control condensin activity. Various parameters can be tested, including the number of active condensins, or the rate that condensin extrudes loop. Direct comparisons between simulation results and biological observations can be applied to identify the exact influence caused by the depletion of condensins, from which an improved understanding of the mechanistic function of condensins can be obtained.

During monitoring the dynamics of structural proteins, we discover that the tension-dependent extrusion mechanism of condensins primarily contribute to the compacting effect. (Ganji et al., 2018) reported their experimental results focusing on the motion of condensin. The loop extrusion mechanism was developed after observing a gradually expanding loop structure on a chromatin formed by a condensin complex in salt water environment provided ATP. Moreover, they observed a decaying extrusion rate as the loop develops, which inspired the hypothesis that the extrusion rate scales inversely to the tension along the chromosome. The tension-dependent extrusion rate mechanism has been embedded in our simulation. After monitoring the local tension, we found that this tension-dependent extrusion rate mechanism

governs the compactness of the chromosome, particularly in the dicentric plasmid chromosome model where the geometry is strictly confined by the tethering centromeres. Methods have been developed to monitor the local tension of chromosomes in time. Comparing to the spatial distribution of condensins and histones, we found a major dependence between local tension and the attachment of proteins. The low tension on a loose chromosome chain allows condensins to extrude loops at a high rate. On the other hand, when large loops already exist, the greater tension localizes on the main chromosome chain and prevents the condensins from traversing rapidly. In addition, histones build tension in the system, and the tension-dependent extrusion activity becomes more sensitive. This automatic adjustment guarantees the existence of large loops, which is the direct causation of condensed structure when exposed to histones.

*Hierarchical loop structures affect polymer mobility*

Polymer stiffness will affect polymer motion. The stiffness of polymers is usually measured by persistence length $L_p$. Rephrasing from Rubinsteins and Colby, the persistence length describes the length scale at which the correlation of polymer tangent vectors decay. In general, greater $L_p$ corresponds to a stiffer polymer. Naked DNA has a persistence length of 50 $nm$. In other words, the motion of DNA segments 50 $nm$ contour length apart would still be correlated.

Studies on traditional polymer models reveals the relationship between the stiffness and the mobility of a polymer. The mobility is measured by mean squared displacement (MSD), which measures the spatial extent of the motion. Polymers with greater mobility exhibit greater MSD. In general, established research states that the MSD of a polymer will increase with growing $L_p$. This can be explained from the perspective of the spring-like characteristics of chromatin (Bloom, 2008).

Derived from the conclusion in polymer physics, there is one hypothesis that the stiffer a naked polymer is, the higher the mobility of the chromosome. However, here we argue that with the existence of geometrical confinement and hierarchical loop structures, the monotonicity vanishes. Hence, we cannot simply deduce the relationship between the stiffness of raw DNA fiber and the mobility of the chromosome in the complex cellular milieu. We support our conclusion by conducting numerical experiments varying the stiffness of the chromatin fiber from $L_p = 5\ nm$ to $L_p = 500\ nm$. We will show that with hierarchical loop structures, the MSD of the chromosome does not increase monotonically as $L_p$ grows. It has been shown in Rouse simulation models that MSD decreases with increasing $L_p$ at short timescales, and increases with increasing $L_p$ at long timescales (Faller & Müller-Plathe, 2001). However, this work only examined chains over a persistence length range of 1-5 monomer diameters.

**Methods**

*Experimental Methods*

To generate strain JLY1056 with *ycg1-2* mutation, first the DNA strain T3000 (Dewar et al., 2004) was cured of the original pT431 strain by growth on YPG and plating on 5 FOA plates. The resultant yeast was transformed with SPC42-mCherry to replace the original Venus tag to generate JLY1038. The JLY1056 strain was generated by transforming JLY1038 with the plasmid p290, kindly provided by Damien D'Amours (University of Ottawa), after digestion with XhoI and XbaI to introduce the ycg1-2 allele. JLY1056 colonies were screened for temperature sensitivity to 42°C.

Mutant SPT10Δ was obtained based on the same carrier dicentric plasmid chromosome pT431. The strain LVY1003 was generated by generating a null allele of SPT10 by removal of the open reading frame using the primers

GTGTACCGATCAAGAACAACTCTAGACTTCCGCCAAAGTGATTATCAACAAAcgga tccccgggttaattaa and GTACAAACTTTATAGTTTCTAGGGTTGGTGATGTGACCGTCTCTGGCAGAGgaattcg agctcgtttaaac. The lowercase letters have homology to pFA6a-KAN, which was used to generate the knockout fragment. The dicentric plasmid, pT431 isolate 2, was transformed into the SPT10::KAN and screened using the GFP signal. All strains were grown in liquid media 24°C.

All biological experiments were performed using at least three different liquid yeast cultures grown on separate days. All strains were grown to midlogarithmic phase before imaging. JLY1056 was grown at 37°C for 3 h before imaging for restrictive temperature imaging, all other strains were grown at 24°C for 4 h before imaging. Strains were grown in YCAT-galactose media with 0.5 mg/ml additional adenine and methionine. Four hours before imaging, cells with the dicentric plasmid pT431 were washed and incubated in SG (synthetic-containing galactose) media lacking methionine with 0.5 mg/ml additional adenine to induce the recombinase to excise the ARS sequence of the plasmid. Cells were then resuspended in YPD for 20 min to repress transcription from the GAL1 promoter and activate the second, conditional centromere (Dewar et al., 2004). Cells were then washed and resuspended in YC-complete media with 2% filter sterile glucose.

Strains were imaged on untreated glass coverslips. Time lapses were performed at room temperature (25°C) using an Eclipse Ti wide-field inverted microscope (Nikon) with a 100× Apo TIRF 1.49 NA objective (Nikon) and Clara charge-coupled device camera (Andor) using Nikon NIS Elements imaging software (Nikon). At each 30-second interval a seven-step Z-stack of 300-nm step size was acquired in the GFP, RFP, and Trans channels a duration of 20 min.

Figure 3.1 shows sampled experimental observations in the GFP, RFP and Trans channels for various experimental groups. The GFP images are of the plasmid label. The RFP images are of the spindle pole body marker SPC42-mCherry. The Trans images are of the yeast themselves using DIC (differential interference contrast) imaging.

Time lapse images of strains containing dicentric plasmids were analyzed using a custom MATLAB GUI, population_GUI_v1_2.m. The GUI allows users to select plasmid signals in the GFP channel, select a background region in the GFP channel for background subtraction, generate and alter a binary mask based on background subtraction of plasmid signal, and select spindle pole bodies in the RFP channel. Only cells containing a single plasmid signal were analyzed. The data for each plasmid signal were saved in a .MAT file for summary analysis of plasmid signal length.

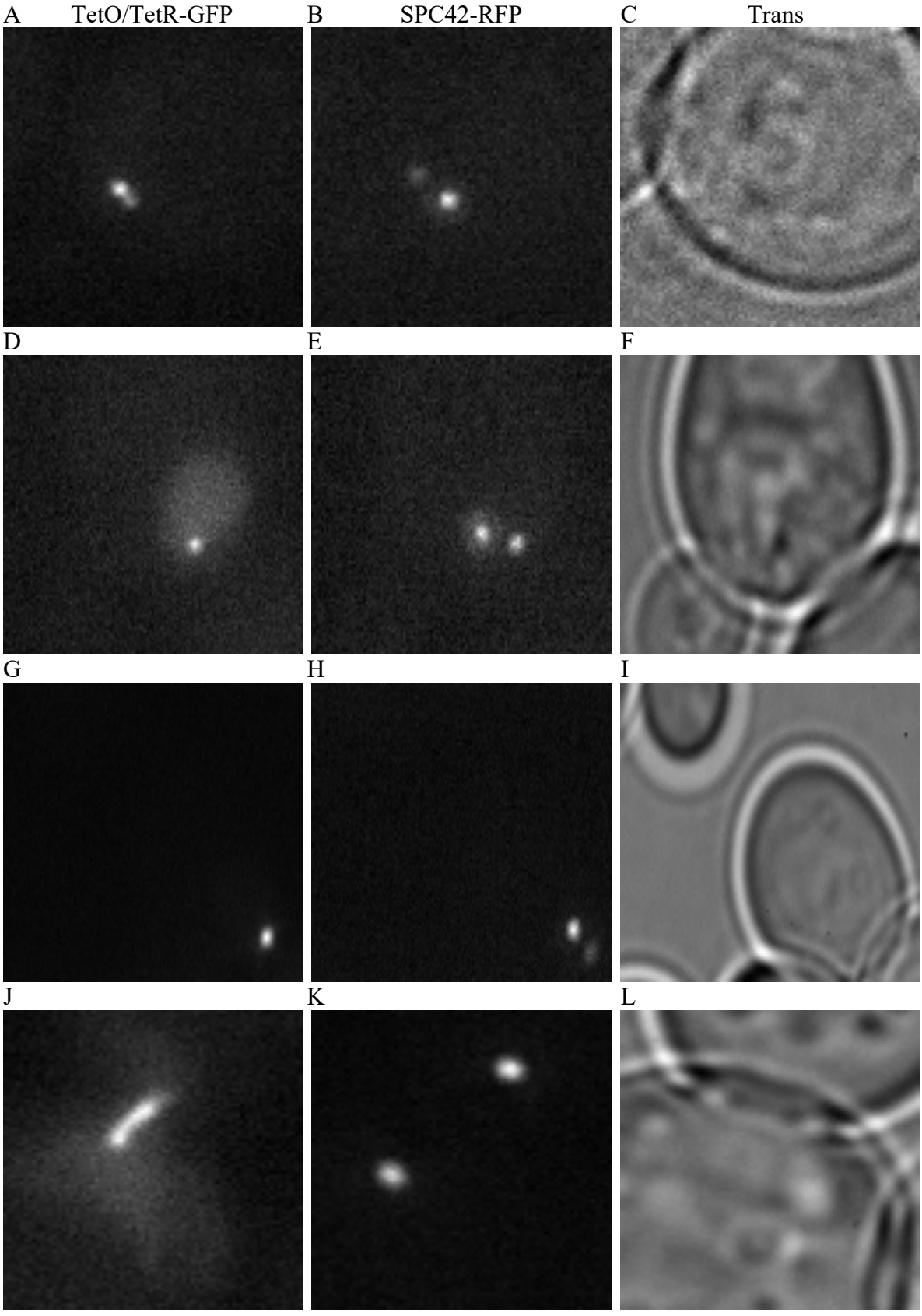

A TetO/TetR-GFP
B SPC42-RFP
C Trans
D
E
F
G
H
I
J
K
L

*Figure 3.1: Images of experimental observations in vivo*
Each column displays observational snapshots from separate channels. From left to right: TetO/TetR-GFP, SPC42-RFP, Trans. Each row displays images obtained from distinct experimental group. From top to bottom: WT 24℃, SPT10Δ 24℃, *ycg*1-2 permissive 24℃, *ycg*1-2 restrictive 37℃.

*Numerical methods simulating biological mutants*

To simulate the impact of SPT10Δ leading to depletion of histone concentration, the histone module is modified. Different methods have been tested to simulate the actual drop in histone concentration. Tested methods include reducing the total number of histones available in the environment; assigning greater proportion of inactive time in one histone period; making histones fragile so that histones detach from the chromosome under tension.

To simulate the varying behavior *ycg*1-2 temperature-sensitive mutant exhibits in permissive condition, we focus on tuning parameters corresponding to condensins while maintaining the other conditions. Likewise, multiple methods have been tested to simulate the inactive condensins, such as the drop in condensin number, malfunctioning condensins by decreasing averaged loop extrusion rate, or tuning the extrusion decay function against tension.

The simulated data were processed through the same visualization pipeline for which microscopic images are processed. Fluorescent signals were simulated, following by a similar approach introduced in Chapter 2 to acquire the signal lengths. The signal lengths can be directly compared to biological data by comparing statistics and distributions. For detailed pipeline to acquire simulated signal from numerical observations, refer to Appendix 1.

During further analysis, we focus on condensins' tension-dependent extrusion rate and determine this mechanism is an important feature in hierarchical loop structures. Hence, we develop methods to analyze tension along the chromosome, and learn the response of condensins in dynamics. The tension can be represented by the length of connecting springs. As tension is built along the polymer chain, adjacent beads will gain local tension, leading to farther distance to balance the tension.

Moreover, we discover methods to characterize the distribution of condensins, and explore the relationship between the distribution of condensins and tension. The attachment

sites of each condensin complex can be set as output of the simulation program, along with the coordinates of the polymer chain at each time step. This relative positional information can be used to obtain the coordinate of each condensin in time and the loop size. However, due to the fact that the output timestep is set to be 10 $\mu s$, which outlasts the averaged lag time between a condensin executing one extruding action, it is impossible to infer the precise extrusion rate backward from geometrical data. One can consider adding temporary extrusion rate as another output feature to monitor the instant extrusion rate.

*Simulations varying polymer stiffness and measuring chromosomal mobility.*

The variation of polymer stiffness is achieved by tuning the parameter "Young's modulus" in Table 1.1. The persistence length $L_p$ scales linearly with Young's modulus governed by the following formula:

$$L_p = \frac{B_s}{k_B T}, \tag{3.1}$$

where $B_s$ is the bending stiffness. In simulations, the bending stiffness of the polymer is derived as a rigid, uniform rod, which results in the following equations

$$B_s = EI, \tag{3.2}$$

where $I = \frac{\pi r^4}{4}$ is the moment of area of the beam with circular cross-section. By default, Young's modulus was selected as 2 GPa to guarantee the polymer stiffness comparable to real pure DNA, 50 *nm*. Hence, scaling Young's modulus from 0.2 to 20 GPa result in a sweep in $L_p$ from 5 nm to 500 nm.

A measurement that is commonly used to inform the mobility of a polymer is mean squared displacement (MSD). MSD describes the diffusive exploration of a polymer locus, calculated as

$$MSD(\tau) = \langle \big(\boldsymbol{r}(t+\tau) - \boldsymbol{r}(t)\big)^2 \rangle, \tag{3.3}$$

where $\tau$ is the lag time, and $\boldsymbol{r}(t)$ represents the position of the polymer at time *t*. In simulated systems, we compute the ensemble-averaged MSD over all fluorescent beads. The discretized formula is given by:

$$MSD(\tau) = \frac{1}{N}\sum_{i\ labeled}\left(\boldsymbol{r_i}(t+\tau) - \boldsymbol{r_i}(t)\right)^2. \quad (3.4)$$

In our system, strong geometrical confinement exists, including tethering and regulation from structural proteins. The confinements lead to sub-diffusive motion, which can be characterized by the corresponding equation in 3D

$$MSD(\tau) = 6D\tau^{\alpha}, \quad (3.5)$$

with a scaling exponent $\alpha < 1$ for restricted lagtimes $\tau$. The pre-factor $D$ is a generalization of the diffusion coefficient with units of $\frac{L^2}{t}$ when $\alpha = 1$. As $\tau$ increases, the scaling exponent $\alpha$ varies and eventually the MSD curve will reach a plateau due to spatial confinement. The MSD value at plateau can be considered as the metric to quantify the motion of the polymer.

We extract positions of the fluorescent beads in the comparison groups with unique $L_p$, calculate the MSDs of the fluorescent beads, and then plot the MSD curves versus $\tau$. The fluorescent portion locates identically across all simulation groups to exclude potential bias. Focusing on the plateau value, we will obtain a measurement for the mobility or motility of the polymer. Furthermore, two sets of these numerical experiments are conducted. One with the raw polymer chain model and another one with DNA-binding proteins included, in order to learn how the hierarchical loop structures modify the dynamical properties of polymers.

We also focus on the dynamics of the proteins while interacting with DNA, especially the dynamics of condensins. These analyses are proven valuable, since the behavior of condensins on either low $L_p$ strain or high $L_p$ strain are significantly different. We exhibit the distinction by plotting the histogram of condensin loop lengths over the entire simulation.

Additionally, we carried out numerical simulations of heterogeneous chromosomes, which obtains distinct properties in different regions. These simulations are implemented by specifying parametrical settings for each bead. The object-oriented design individualizes the connecting springs, which allows regional modifications of all components, including stiffness. In simulations, a heterogeneous plasmid chromosome was simulated, which consists of two distinct segments. The segment from Bead 69 to Bead 262 possess one choice of stiffness, and the rest of the chromosome obtains another. Notice that the plasmid chromosome obtains a circular geometry with two tethered sites located at Bead 0 and Bead 138, the split of the heterogeneous chain is topologically symmetric. The symmetry remains to the labeled region, since the labeled region starts from Bead 161 and end at Bead 352. These 192 beads are split into two parts with comparable sizes.

## Results

### *In a condensin-concentrated environment, histone enrichment leads to stable DNA structure*

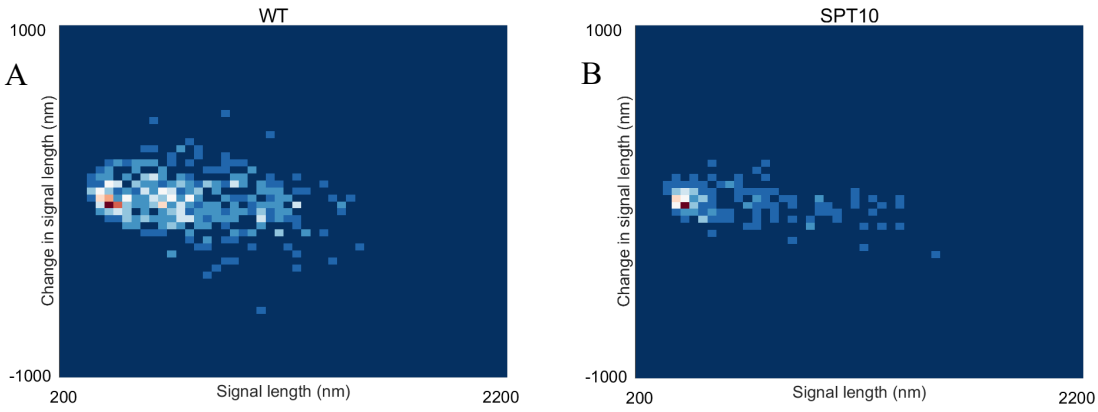
In experiments, the signal length histogram of SPT-10 experimental group is similar to WT comparison group, except for the more extreme cases that appear in SPT-10 control group, on both small side and large side (Figure 3.2). In other words, the histone-mutated environment fails to make significant difference to the chromosome structure, on the contrary to what researchers expected that a lack of histones will induce less condensed chromosome. Alternatively, it causes more extremely condensed cases of less than 500 nm and more extremely stretched cases of more than 1200 nm. Qualitative features in support can also be found in violin plots and a density heatmap.

To explore in detail what leads to these consequences, we apply the simulation to reproduce the experiment. We apply the simulation with parameters that fit the WT experimental data as our comparison group. For controlled group, to simulate SPT10Δ

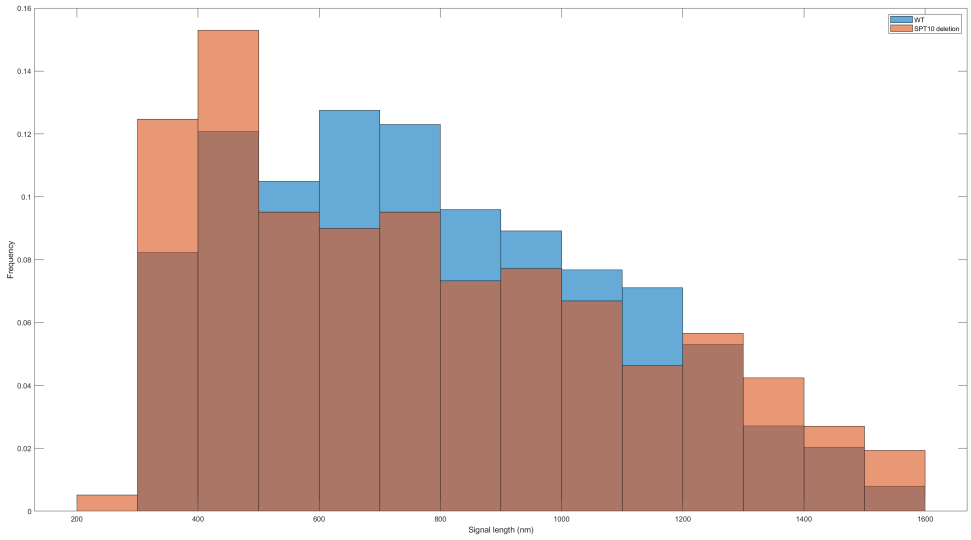
environment, we modify the parameters corresponding to histones to partially mute the histones.

The most straight-forward modification is to reduce the number of histones enriched in the environment. In simulation it can be achieved by lengthening the duration for histones to detach from the chromosome chain in a histone period. When we reduce the active histone to 50% of the standard level, we observe similar features in plasmid signal length distribution as we observe in experiment (Figure 3.3). Compared to the simulation control group, the experiment group obtains more samples on extremely small side, similar to experimental conclusion.

In order to explain this, we closely analyze the condensin behavior. In histone-enriched environment, condensins turn out to be less active. They tend to stay in the same position for a longer duration, hence they fail to construct larger loops. On the contrary, in histone-muted environment condensins are more actively extruding loops. Notice that our previous studies have showed the structure without any histones (Figure 2.4). The signals widely spread as the consequence of dynamic extrusion, but large loops fail to collapse in size, leading to the unshifted large averaged signal size. This implies that in a condensin-enriched environment, concentrating histones does not condense the structure monotonically. The hierarchical loop structures interact with one another, resulting in complicated effects on chromosome organization. In conclusion, histones effectively condense large loop structures formed by other proteins (including condensins). Meanwhile, histones build local tension while attaching to the DNA, which in turn limits the motility of condensins. These conflicting mechanisms balance, leading to an optimal organization that possesses both compactness and mobility.


A
WT
1000
-1000
Change in signal length (nm)
200
Signal length (nm)
2200
B
SPT10
1000
-1000
Change in signal length (nm)
200
Signal length (nm)
2200

C


WT
SPT10 deletion
Frequency
Signal length (nm)

D

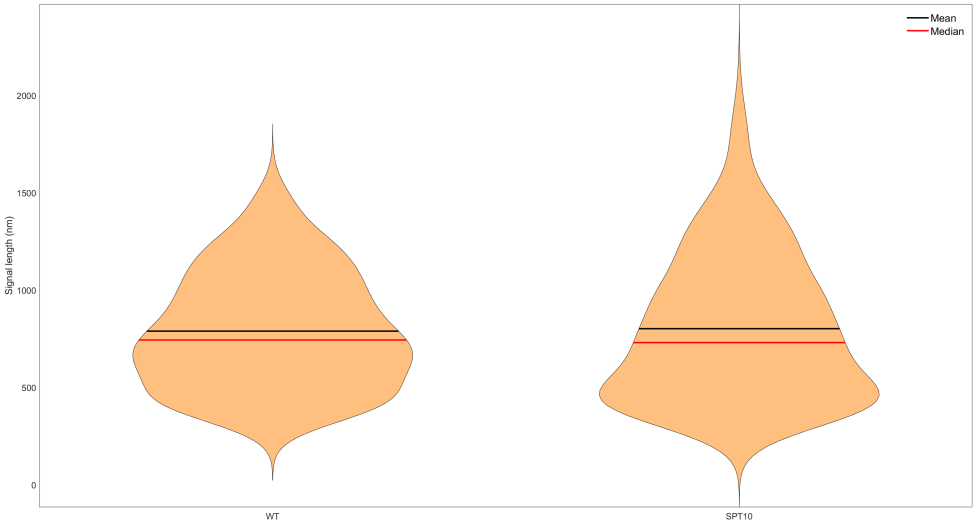
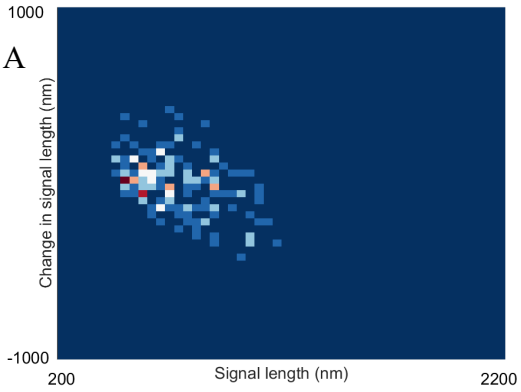
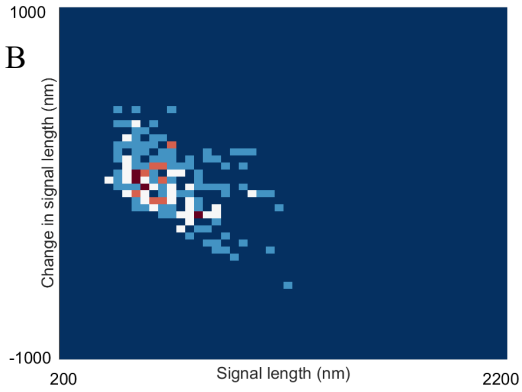

Mean
Median
Signal length (nm)
WT
SPT10

*Figure 3.2: Experimental Report on WT vs. SPT10Δ plasmid signals*
Statistical report on WT plasmid signals compared with SPT10Δ mutants produced in experiments. (A) Density heatmap of WT plasmid signals. (B) Density heatmap of SPT10Δ mutants signals. (C) Combined signal length histogram of WT and SPT10Δ plasmids. (D) Combined signal length violin plot of WT and SPT10Δ plasmids.

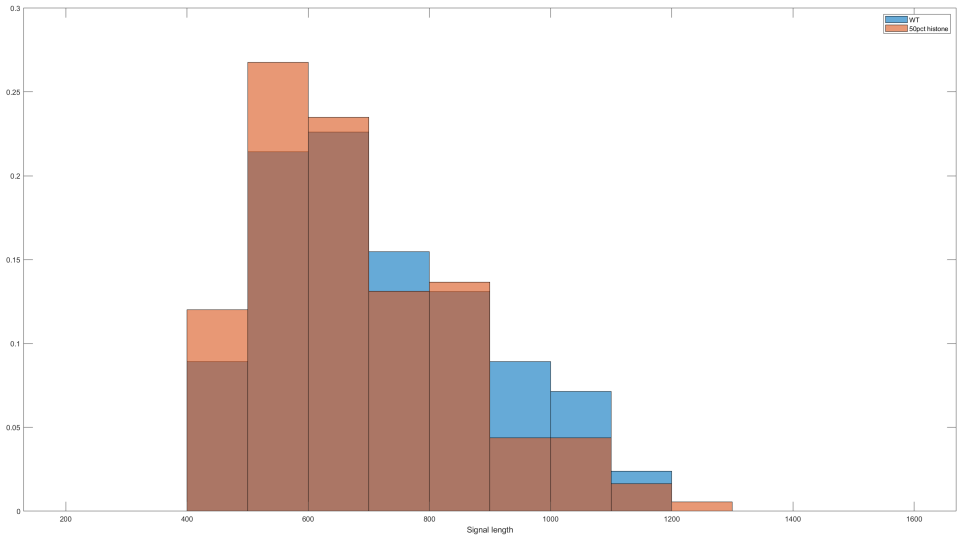
A


1000
-1000
Change in signal length (nm)
200
Signal length (nm)
2200

B


1000
-1000
Change in signal length (nm)
200
Signal length (nm)
2200

C


WT
50pct histone
0.3
0.25
0.2
0.15
0.1
0.05
0
200
400
600
800
1000
1200
1400
1600
Signal length

D

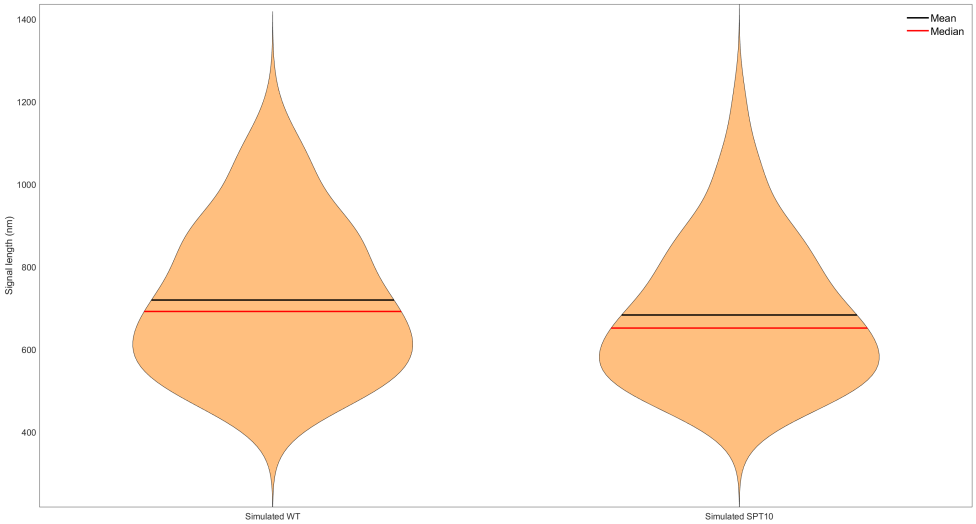

Mean
Median
1400
1200
1000
800
600
400
Signal length (nm)
Simulated WT
Simulated SPT10

*Figure 3.3: Report on simulated WT vs. SPT10Δ plasmid signals*
Statistical report on simulated WT plasmid signals compared with SPT10Δ mutants. In simulated SPT10Δ mutants, the average number of active histones is reduced to 50% of standard level. (A) Density heatmap of WT plasmid signals. (B) Density heatmap of SPT10Δ mutants signals. (C) Combined signal length histogram of WT and SPT10Δ plasmids. (D) Combined signal length violin plot of WT and SPT10Δ plasmids.

*DNA is condensed under the existence of condensin-mediated loops in dicentric plasmid chromosome*

In *ycg*1-2 experiment (Figure 3.4), we observe a larger averaged signal length for *ycg*1-2 restrictive experiment compared to *ycg*1-2 permissive observations, which are regarded as the control group. More observations showed over 800 nm signal length, which was not often observed in the control experiment. A more detailed statistical analysis showed a clear bimodal distribution of signal length. The lower peak is distributed around 400 nm, similar to WT signal length. The larger peak press close to 900 nm, which is similar to raw chain cases without any condensins. The bimodal distribution deviates from the control group with right-skewed distribution.

To simulate *ycg*1-2 restrictive mutant environment which suppresses condensins, the number of active condensins in the simulation was reduced from 6 to 3. These condensins actively extrude loops, but occasionally release the loop and restart its extrusion activity. Initializations corresponding to histones and the DNA chain remain unchanged.

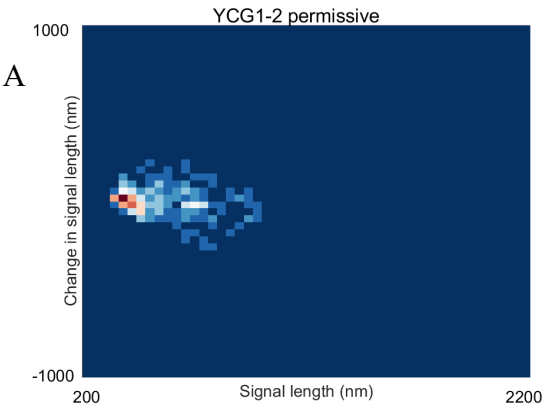
The simulation result shows a similar bimodal property as observed in experimental data (Figure 3.5). The two peaks locate near 400-500 nm and 800-900 nm. The violin plot also displayed additional mode with increased average signal length, leading to pear-shaped plot. Mean and median are between two peaks, and the fact that mean and median shift order implies the change of skewness. Remarkably, if we analyze the heatmap of both experimental *ycg1-2* data and simulated *ycg1-2* data, an interesting fact is that there are two clusters of points, suggesting the possibility that the system oscillates between two phases. One phase with denser structure corresponds to WT cases, obtaining a compact structure of 400 nm. The other phase tends toward cases without condensins, with an averaged 900 nm signal.

A hypothetical interpretation would be that the number of observations where no condensin interacts with the labeled region rises to a significant level, because of the lack of

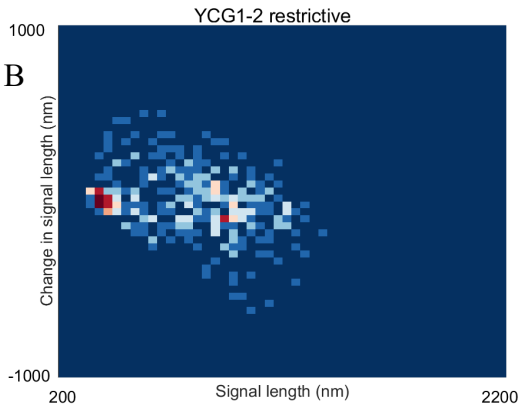
condensins. This leads to two primary phases: when condensins form loops involving the labeled region, the signal regulates similarly to WT signals; when condensins do not interact with the labeled region, the signals retain condensin-excluded properties, showing large signals. Moreover, the reduction of condensins causes lower tension along the chain, which in turn enhances the motility. This is supported by the wider density heatmap in vertical direction (Figure 3.4 B). Hence, the overall behavior of chromosome should be highly oscillatory between the two modes.

The analysis showed that the enlarging loops formed by condensins attain the power to condense the chromosome structure in our model. In addition, recall that in the histone-free simulations, condensins fail to compact the chromosome (Figure 2.4). When histones exist, the large loop structures are further condensed by small local loops, which effectively reduces the volume of the chromosome. This leads to the conclusion that condensins cooperate with histones to condense and compact the chromosome.

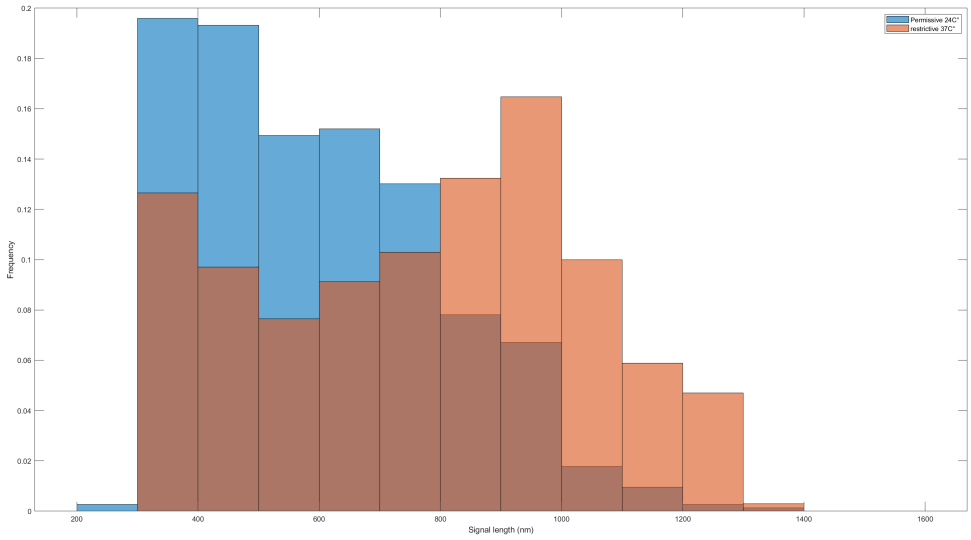
A


YCG1-2 permissive
1000
-1000
Change in signal length (nm)
200
Signal length (nm)
2200

B


YCG1-2 restrictive
1000
-1000
Change in signal length (nm)
200
Signal length (nm)
2200

C


Permissive 24C°
restrictive 37C°
Frequency
Signal length (nm)

D

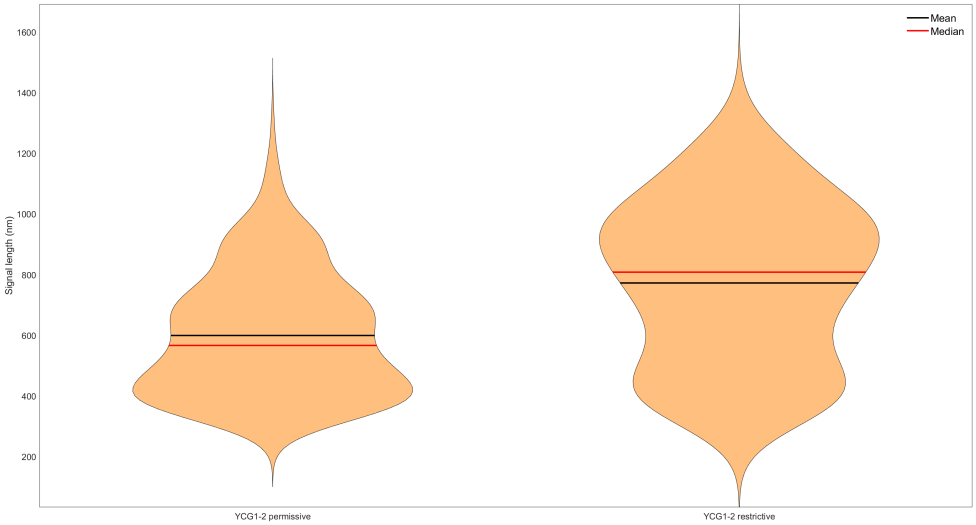
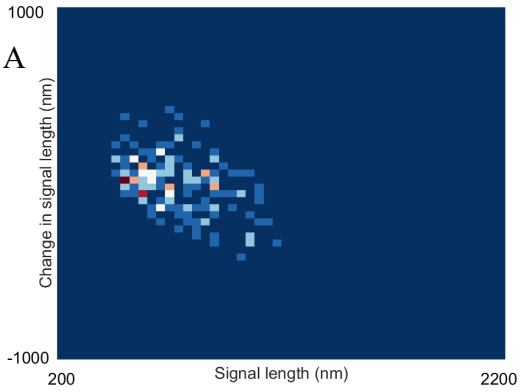
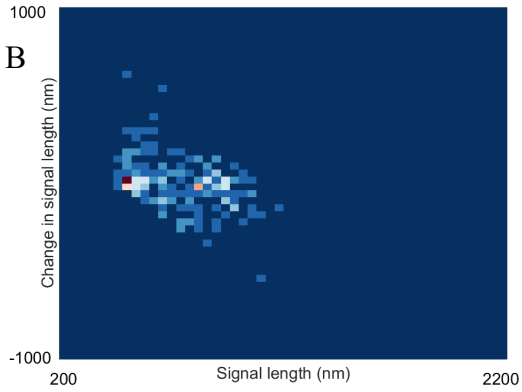

Mean
Median
Signal length (nm)
YCG1-2 permissive
YCG1-2 restrictive

*Figure 3.4: Experimental report on WT vs. ycg1-2 plasmid signals*
Statistical report on WT plasmid signals compared with *ycg*1-2 mutants produced in experiments. (A) Density heatmap of WT plasmid signals. (B) Density heatmap of *ycg*1-2 mutants signals. (C) Combined signal length histogram of WT and *ycg*1-2 plasmids. (D) Combined signal length violin plot of WT and *ycg*1-2 plasmids.

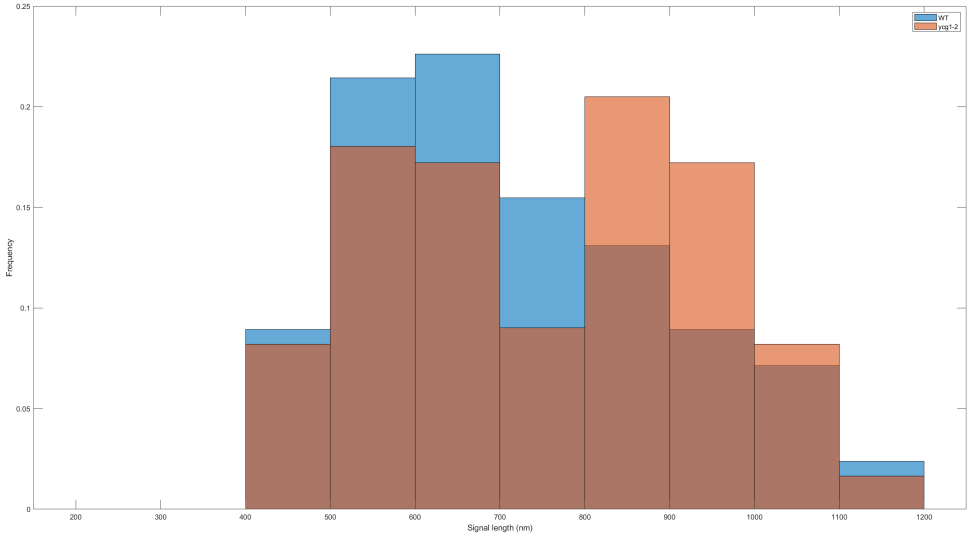
A


1000
Change in signal length (nm)
-1000
200
Signal length (nm)
2200

B


1000
Change in signal length (nm)
-1000
200
Signal length (nm)
2200

C


WT
ycg1-2
Frequency
Signal length (nm)

D

xx

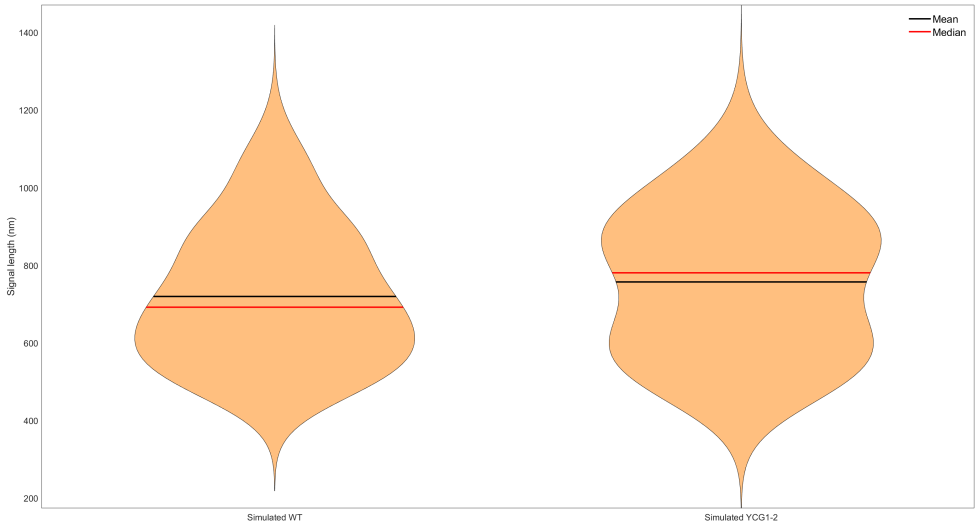

Mean
Median
Signal length (nm)
Simulated WT
Simulated YCG1-2

*Figure 3.5: Report on simulated WT vs. ycg1-2 plasmid signals*
Statistical report on simulated WT plasmid signals compared with *ycg*1-2 mutants. In simulated *ycg*1-2 mutants, the number of active condensins is reduced from 6 to 3. (A) Density heatmap of WT plasmid signals. (B) Density heatmap of *ycg*1-2 mutants signals. (C) Combined signal length histogram of WT and *ycg*1-2 plasmids. (D) Combined signal length violin plot of WT and *ycg*1-2 plasmids.

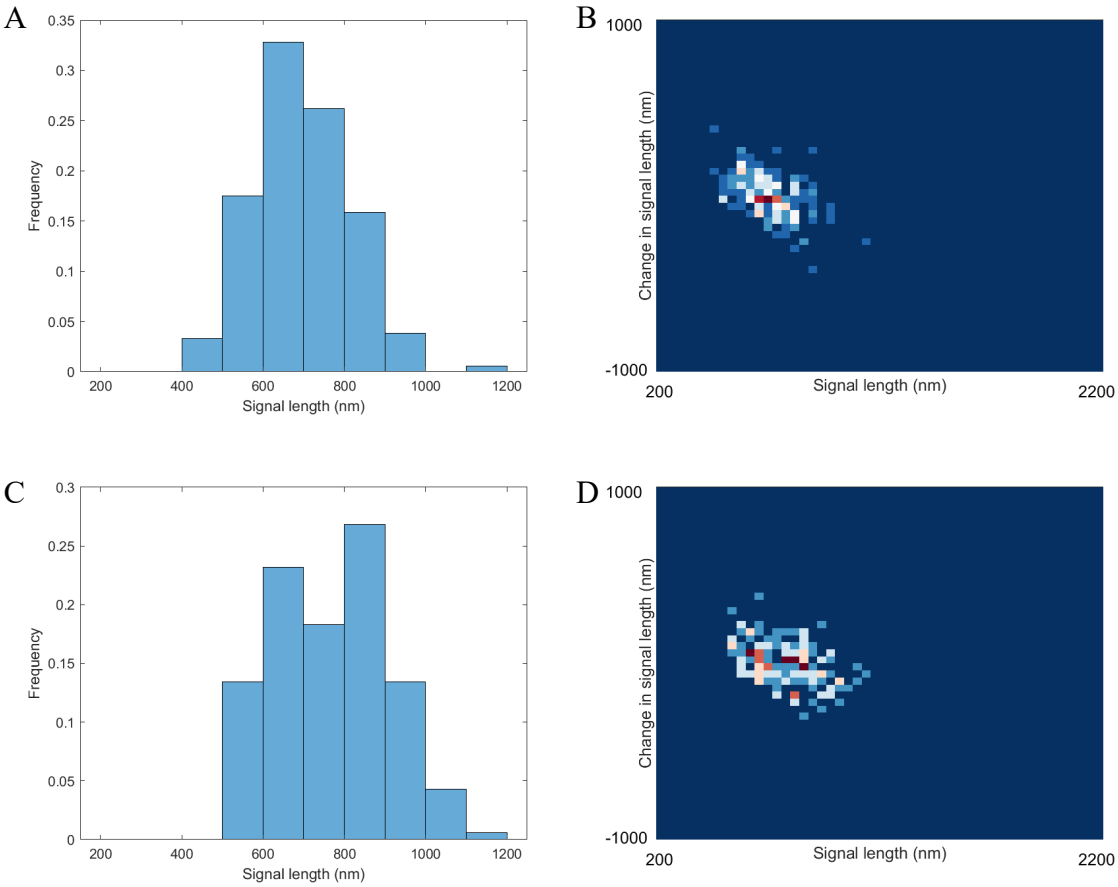
*Tension-dependent extrusion behavior for condensins benefits formation of effective loops.*

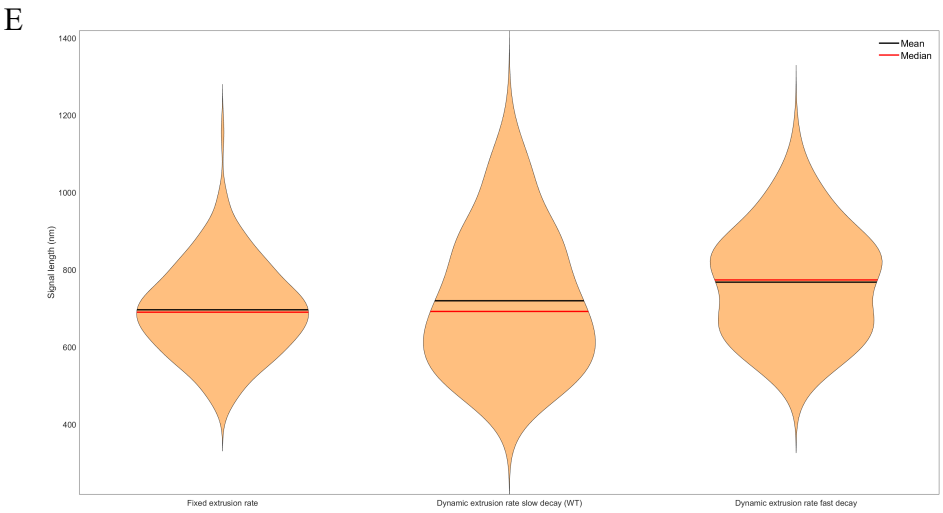
An alternative modification we tested to simulate the depletion of condensin activity is varying the tension-dependent extrusion decay rate. As is reported by Ganji et al., a condensin extrudes loops in an extension-dependent manner in a flow-stretched separate environment. This observation is mechanistically interpreted as tension-dependent extrusion behavior. Similar to other DNA-binding proteins, condensins are able to detect regional tension along the DNA chain at the attachment site and alter their behavior accordingly. We simulate this mechanism by introducing an embedded extrusion rate decaying function for condensins.

The tension-dependent extrusion mechanism is tested in simulation (Figure 3.6). When the extrusion rate remains fixed and independent of tension, condensins keep extruding loops, leading to a more uniform distribution in loop size extruded by condensins (Figure 3.6 F). However, the uniform distributed condensin loop size does not lead to effective chromosomal compactness. The signals obtain Gaussian-like distribution centering at 600 nm ~ 700 nm. Extreme cases are less frequently observed compared to tension-dependent situations.

On the other hand, if we extremize the tension-dependent decay mechanism and make the extrusion activity excessively sensitive to tension, we observe that condensins' ability to extrude large loops degenerates. The concentration in small-size loops displayed in (Figure 3.6 F) showed dysfunction of condensins in extruding large loops. As a result, a bimodal distribution is observed with two centers located at 600 nm and 900 nm.

In conclusion, the tension-dependent extrusion rate is a key mechanism for condensins when regulating chromosomal structures. It enables condensins to form and maintain large loops, which act as the primary factor leading to the compact structure of chromosomes.


A
Frequency
Signal length (nm)
B
Change in signal length (nm)
Signal length (nm)
C
Frequency
Signal length (nm)
D
Change in signal length (nm)
Signal length (nm)


E
Mean
Median
Signal length (nm)
Fixed extrusion rate
Dynamic extrusion rate slow decay (WT)
Dynamic extrusion rate fast decay

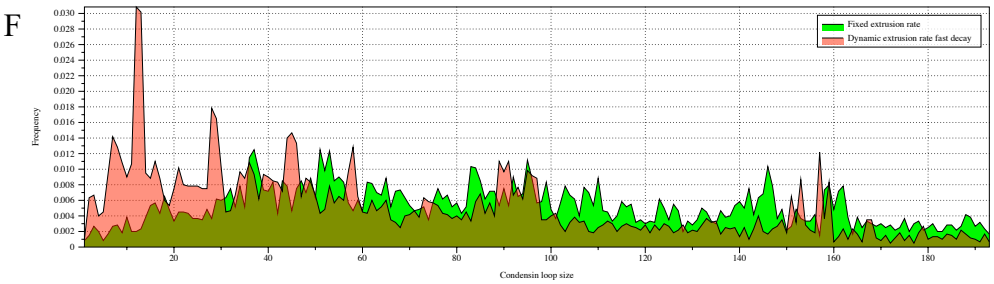

F
Fixed extrusion rate
Dynamic extrusion rate fast decay
Frequency
Condensin loop size

*Figure 3.6: Comparison plots varying condensins' tension-dependent extrusion behavior*
(A)(B) Signal length histogram and density heatmap for simulation results where condensins extrude loops at a constant rate. (C)(D) Signal length histogram and density heatmap for simulation results where condensins extrude loops at a tension-dependent rate. The extrusion rate decays sharper against increasing tension, compared to normal rate of decay (WT). (E) Violin plot of signal lengths as a function of the rate of decay of loop extrusion against tension. (F) Condensin loop size histogram sampled in one run, as a function of the rate of decay of loop extrusion against tension. All simulations obtain the same parametrization, including the default extrusion rate of condensins.

*Hierarchical loop structures modify chromatin stiffness*

In studies of polymer physics, the polymer stiffness is strongly correlated with the mobility of the polymer due to thermal fluctuations. The mobility of a polymer is determined by its MSD, which describes the space the polymer explores in a restricted time interval. Studies have shown that for arbitrary polymer models, the MSD will increase with increasing $L_p$. From the perspective of the spring-like characteristics of chromatin where we have (Bloom, 2008)

$$k_s = \frac{k_B T}{L_p^2}. \tag{3.6}$$

The spring constant increases with increasing temperature $T$ and decreases with increasing $L_p$. This reflects the fewer entropic states that chains with greater persistence lengths can adopt. A weaker spring will exhibit more variance in fluctuations and may therefore exhibit greater motion.

This aligns with our chromosome dynamics simulation results in the absence of histone compaction and condensins (Figure 3.7 A, solid lines). The solid lines represent the MSD plot of raw polymers without any structural proteins as a function of lag time $\tau$. Notice the monotonicity of MSD with increasing $L_p$. This shows consistency with theoretical analysis.

However, the monotonicity breaks down if histones *and* condensins are included (Figure 3.7 A, dashed lines). As $L_p$ increases, the MSD increases at first for flexible chains, but decreases for stiff systems with $L_p$ in the 200-500 nm range. This implies that the loop structures affect the mobility of a polymer and modify its fundamental structural properties. While chromatin changes that increase or decrease $L_p$ will clearly manifest as changes in MSD, the relationship between these parameters cannot be simply deduced from experimental observations in the complex cellular milieu.

We explore the causation by examining the behavior of condensins in the systems. Figure 3.7 B shows sampled condensin loop size histogram during one simulation. Besides the loop size, the motility of condensins can also be interred from this figure. For highly flexible chains ($L_p = 5\ nm$), the loops are highly dynamic and exhibit a broad distribution. For stiffer chains ($L_p = 500\ nm$) the loops cluster into discrete size classes (number of beads/loop).

Similar conclusions can also be observed in the contact frequency map (Figure 3.8). Condensins result in contacts between separated segments of DNA by forming crosslinks between them. Histones, on the other hand, lead to local contacts, which is represented by increased contacts in regions near the diagonal. In flexible chains, the average contacts were evenly increased, with few strong spots. However, in stiff chains, condensins lead to strong contacts between settled regions. This supports the statement that while histones are enriched, condensins are more static in stiffer chains. They prefer to form large, stationary loops. Meanwhile, condensins in a flexible chain are more dynamic. This property is not only determined by the tension-dependence of condensins, but also affected by the geometrical restriction prescribed in plasmid system.

Presumably, this phase separation of the loop structure in systems with varying stiffness causes the property change in chromatin mobility. In stiff strains, large, static loops are further compacted by histones. As a result, the chromosome chain is highly condensed with very limited fluctuations in space. The synergetic mechanism of hierarchical loops formed by histones and condensins leads to the sharp decrease in MSD for stiffer chains.

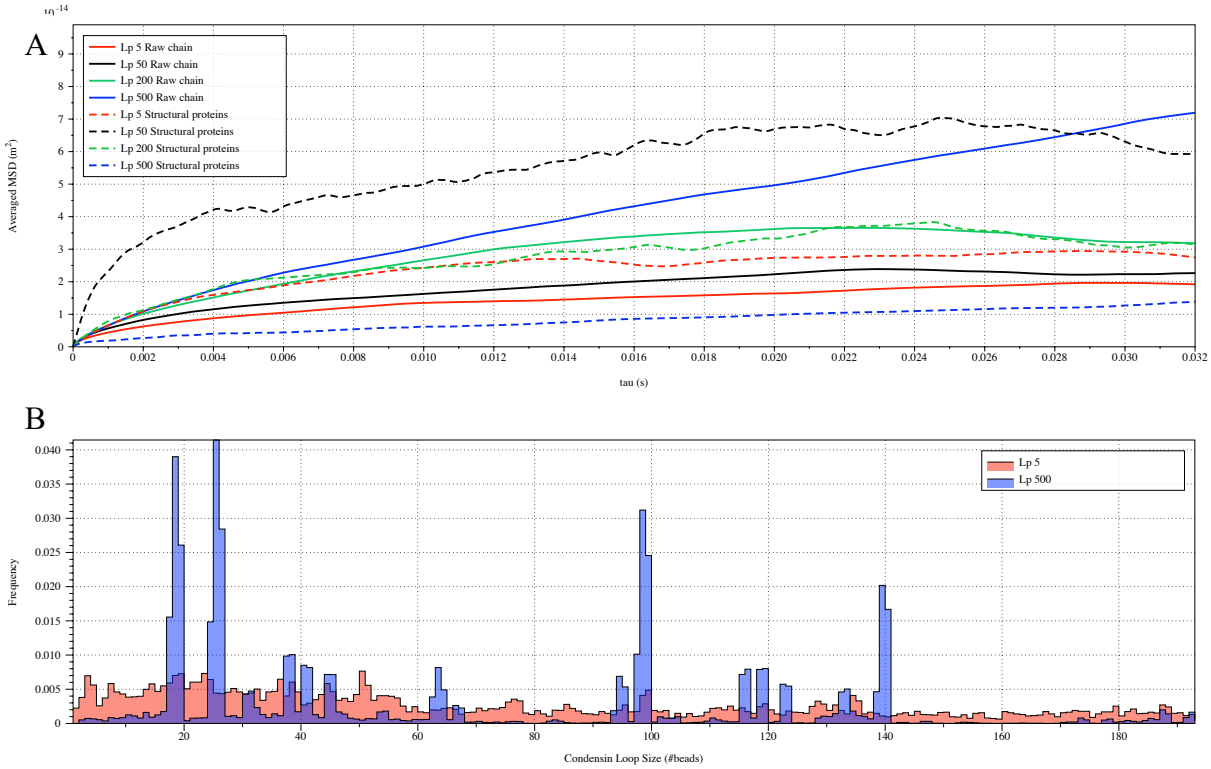



*Figure 3.7: Varying stiffness lead to unexpected results with hierarchical loop structures* (A) Simulated MSD plot as a function of $\tau$. $L_p$ varies from 5 *nm* to 500 *nm*. Solid lines correspond to simulations in the absence of any proteins. Dashed lines demonstrate statistics of simulation results with both histones and condensins. (B) Cumulative distribution in the sizes of loops for condensin loops, compared between highly flexible chains in red ($L_p = 5\ nm$) and stiffer chains in blue ($L_p = 500\ nm$).

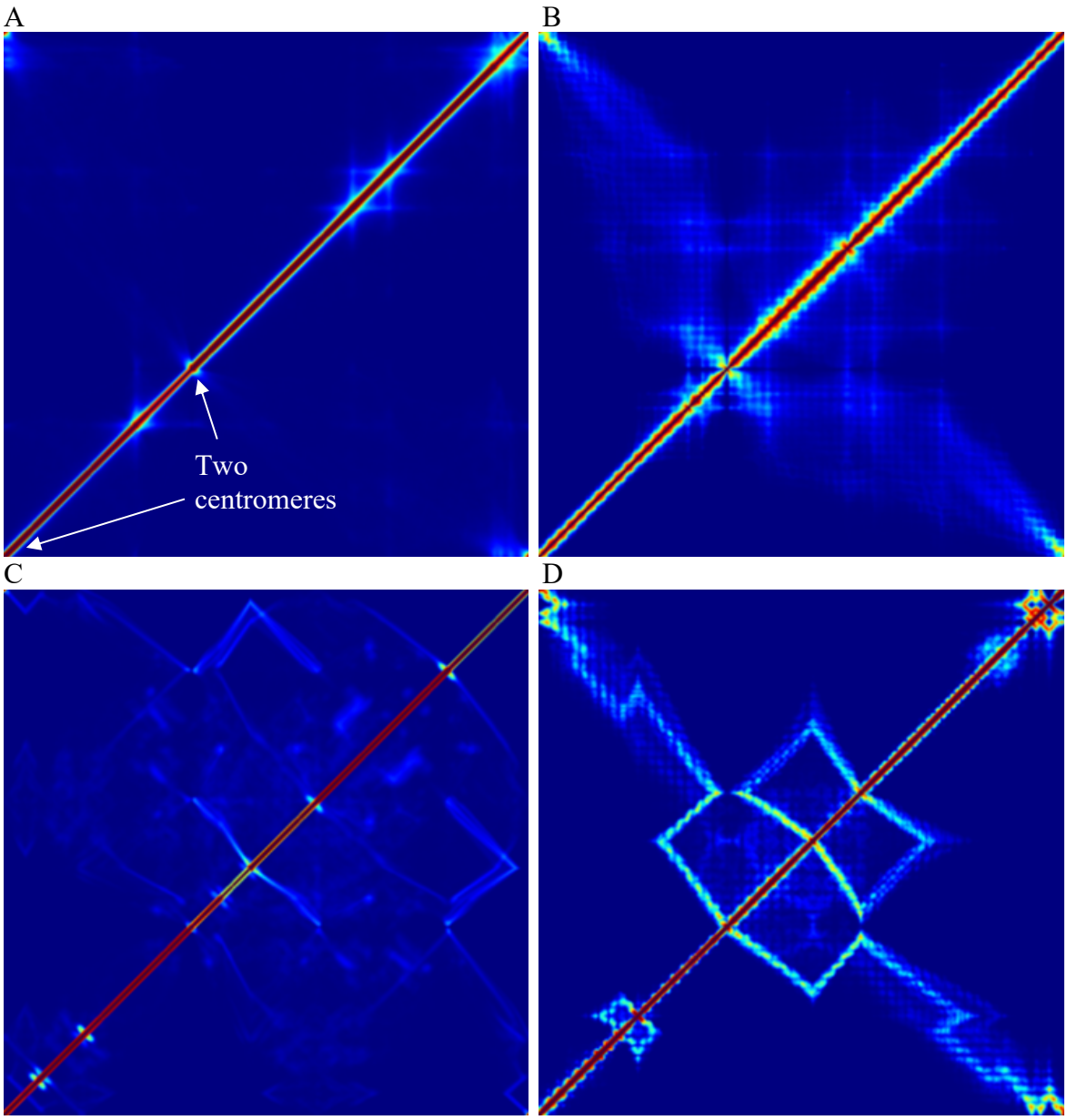



*Figure 3.8: Pairwise contact frequency map of simulated plasmids*
Pairwise contact frequency map of simulated plasmid chromosome with varying conditions. Contacts are measured by counting frequencies where two beads are spatially close. (A) Condensins $L_p = 5\ nm$. (B) Condensins and histones $L_p = 5\ nm$. (C) Condensins $L_p = 500\ nm$. (D) Condensins and histones $L_p = 500\ nm$. Red region represents high pairwise contact frequency in contact map. Zero values in diagonal due to ignorance of self-contacts. Beads corresponding to two centromeres are labeled in Figure A.

Alternative numerical experiments were designed to introduce heterogeneous segments on one chromosome chain. In the experiment, the circular plasmid chromosome is evenly split into two regions carrying distinct polymer properties. Meanwhile we also introduce proteins including histones and condensins in the simulation. Then the MSD was analyzed over the fluorescent region. Provided that the fluorescent region is also evenly divided into the two heterogeneous segments, the bias of comparison in regional MSD is minimized.

Similar to the experiments conducted on independent chains, monotonicity also vanishes on the heterogeneous chains (Figure 3.9). The MSD plateau of regions with $L_p = 50\ nm$ dominates both plateaus of alternative regions $L_p = 5\ nm$ and $L_p = 200\ nm$.

This supports the conclusion that hierarchical loops modify the mobility of the chromosome. For stiff regions with $L_p > 50\ nm$, the dynamics of condensins are significantly restricted. For heterogeneous chromosomes, this will cause a concentration of condensins in the stiff region in time. With the compacting effect of concentrated hierarchical loops, the constraints in diffusion dominate the extra mobility caused by greater polymer stiffness. As a result, the MSD of a stiffer region with $L_p = 200\ nm$ drops compared to the flexible region with $L_p = 50\ nm$. In the other experiment with less overall stiffness (half $L_p$ 5 $nm$, half $L_p$ 50 $nm$), the effect of hierarchical loops is not significant enough to overcome the effect of amplified stiffness, and thus the MSD still increases with increasing $L_p$.

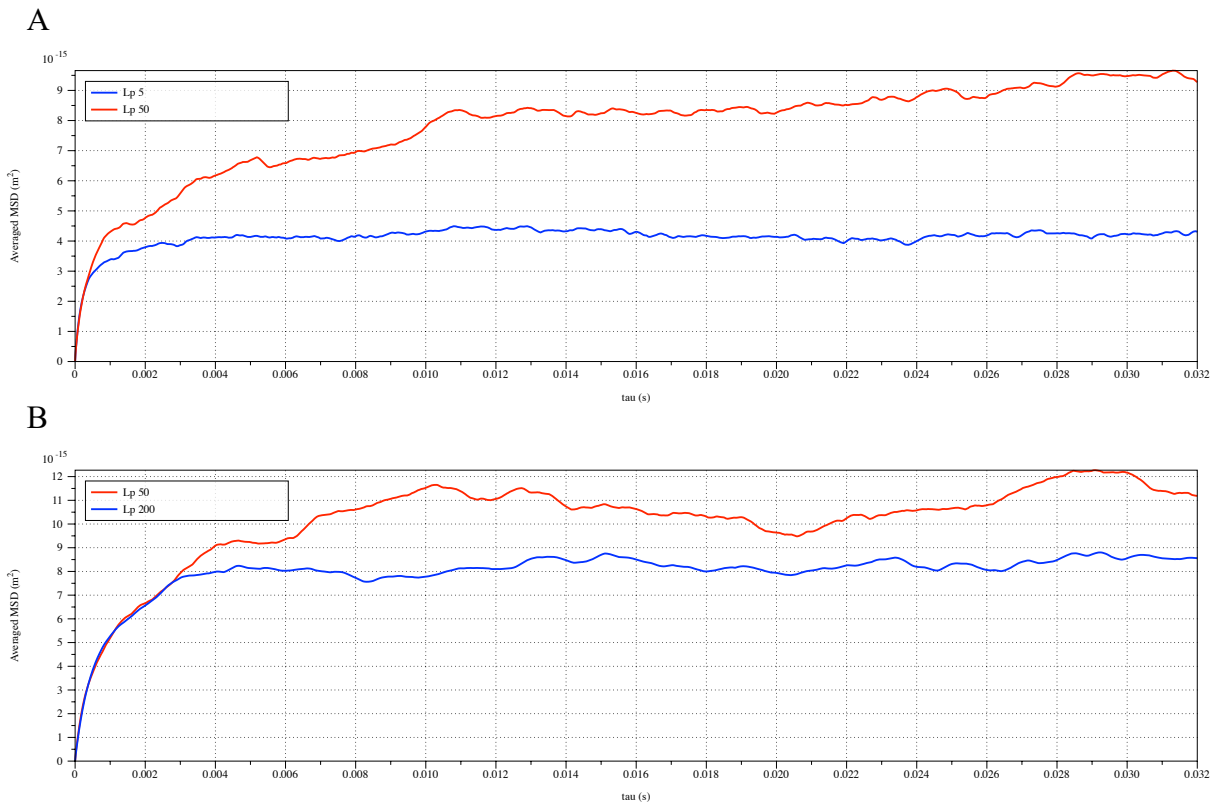



*Figure 3.9: Reginal MSD plots for simulated heterogeneous plasmid chromosomes*
Regional MSD plots over the fluorescent region of simulated heterogeneous plasmid chromosomes. The plasmid is evenly split into two halves of unique stiffness. (A) Half $L_p = 5\,nm$ vs. half $L_p = 50\,nm$. (B) Half $L_p = 50\,nm$ vs. half $L_p = 200\,nm$. The data was selected from samples in evenly split fluorescent region.

## Discussion

*The balance between condensins and histones affects the compaction effect*

Surprisingly, slight variation in numerical properties of the proteins will cause unanticipated effects on the chromosomal organization. We maintain the histone settings and modify the averaged extrusion rate of condensins (The default extrusion rate in tension-free system is increased, while the extrusion decay function remains unchanged). In both cases where the averaged condensin extrusion rate are either increased or decreased, we observe increases in the averaged signal length. This observation does not appear if histones are excluded in the simulation environment.

This result might lead to a conclusion that there is a balance within the hierarchical structures, resulting in a compact chromosome structure. We seek explanations through simulation, and insights are found if the dynamics of condensins are carefully examined. Condensins with low extrusion rate barely function as condensing proteins. The loops generated by condensins are small, resulting in a similar output with no condensins. On the other hand, condensins with high extrusion rate keep extruding loops along the chain, bringing in too much dynamics to the system. The signal length histogram is spread out compared with the experiment with reasonable averaged condensin extrusion rate. Similar result can be observed if the extrusion decay function against tension is modified. In both cases where condensin extrusion rate decays more drastically or mildly against tension along the chain, the chromosomes obtain a looser structure.

Clearly the restriction of geometrical configuration of plasmid chromosome is a significant factor that determines such a consequence. Condensins are able to extrude larger loops on larger chromosomes, which might end up with an alternative optimal choice of

parameters. Hence further exploration on the loop structures in other chromosome systems will be meaningful.

*Tension-dependent extrusion rate guarantees the compactness and mobility of chromosomes.*

The observation of condensins revealed by Ganji et al. suggests three major discoveries. Firstly, the behavior of condensins inclines to an extrusion behavior rather than forming random crosslinks on two DNA loci. Secondly, the extrusion activity is strictly ATP-hydrolysis dependent. Lastly, the extrusion rate decays as the DNA extends relatively, due to the flow-stretched spatial confinement in the experiment. However, discussions remain to explain the underline reason for cause of the decay of the extrusion rate.

In experiment, the DNA obtains specific geometrical structure. A linear piece of $\lambda$-DNA is tethered at two ends with an end-to-end distance smaller than its contour length. Hence as the condensin complex builds up the loop, the DNA gradually gets stretched. A reasonable hypothesis brought up by Ganji's group is that extrusion rates of condensins decay as the tension is built along the attached DNA. This hypothesis is also adopted and embedded in this simulation work. During the exploration to find proper parametric initialization for extrusion decay rate of condensins, we found that the decay of extrusion rate applies a great effect on compacting chromosome structure, and a precise choice of parameters leads to optimal compaction.

When condensins either maintain a constant extrusion rate or decay subtly throughout the simulation process, we observe too much dynamics caused by condensins. The loops created by condensin complexes continuously vary their sizes, resulting in low averaged loop size due to the restricted, circular structure of the plasmid chromosome. On the other hand, in decaying condensin extrusion rate model with extrusion rate inversely proportional to tension (which is selected as simulation for wild type condensin), condensin loops develop in early

stage and preserve a large loop scale as the tension has been built along the chromosome chain. Large loops contribute to compression of the volume in the presence of histones. In other situations where a sharper decaying function is assigned to the extrution rate against tension, large loops are impossible to form in the first place. If condensin extrudes at a rate inversely proportional to square of tension, condensin complexes come easily to a halt when very little tension forms, ending up with small loops. Those small loops basically do not contribute to reducing chromosome volume.

The numerical experiments varying $L_p$ also offer a glimpse into the essence of extrusion decay. The ability for condensins to detect local tension and respond accordingly grants condensin the intelligence to adopt an optimized loop size, which is beneficial in preserving compactness while maintaining motility.

It would be very helpful to conduct biological experiments modifying the stiffness of DNA, similar to what we have done in simulation. There is a hypothesis that biochemical modifications to histones, called post-translational modifications (PMTs), may change the stiffness of chromatin (Bowman & Poirier, 2015). In addition, the PMTs of histone can be regional. Hopefully, there will be experimental results that we can compare our simulation results to.

*Indication of existence of Z-loops*

Figure 3.8 D shows the pairwise contact frequency for stiff polymers with $L_p = 500\ nm$, which is 10 times stiffer than common DNA. However, the simulation is still valuable, for the fact that qualitative conclusions are more visible and are easy to interpret in this system. One major feature the contact map shows is the line segments. A bright polyline connects two centromeres (Bead 0 and Bead 138). Through careful observation, we explain the formation of such a bright polyline by the tightly tethered backbone of the plasmid

chromosome. The loop structures create local substructures, occupying a large portion of the chromosome. The remaining segments are tightly tethered by the centromeres, forming two parallel lines between the poles. Contacts are accounted for two nearby beads along the plasmid backbone. This explains that there are steady contacts between beads which are far apart, and the pairs connects two centromeres.

Besides, there are minor square structures connecting two arbitrary beads. These are evidence for Z-loops. One large loop formed by a condensin lands one of its attachment sites on another loop structure, leading to a double-connection structure (Figure 3.10 A). Under external tethering and protein compacting, this structure can hardly diffuse under thermal fluctuation. Consequently, the existence of Z-loops remarkably reduces the volume of the chromosome.

Not only observed in stiff chromosomes, this structure is commonly observed in other simulated chromosomes as multiple dynamic loops intertwine and interact (Figure 3.10 B). A quantitative method to demonstrate the existence of Z-loops is to plot the projection distance from Centromere 1 as a function of bead index. As is shown in Figure 3.10 C, the curve obtains an increasing trend from the starting centromere (Bead 0) to another centromere (Bead 138), and tends to decrease while circling back. Local zigzags are usually caused by small-scale loops formed by histones. Flat regions or oscillating regions may correspond to large, floppy loops caused by condensins. An inverse trend, such as the emphasized region in Figure 3.10 C, is very likely to be the indication of Z-loops, since Z-loops will cause an inversely aligned region under tension.

The Z-loop structure can be regarded as a higher order loop structure. In terms of function in the nucleus, it efficiently enhances chromosomal compaction. Moreover, the appearance of Z-loops, the three-fold loop structure consisting of two condensin complexes, indicates the existence of various higher-order looping structures involving multiple DNA-

binding proteins. And thus, condensins may achieve chromosomal condensation via various higher-order looping structures.

A recently published work announces that the Z-loop structure cause by condensins has been visualized in microscopy (Kim et al., 2020). The experimental approach was similar to the method Ganji's group adopted: utilizing real-time imaging techniques to observe a tethered, separated, flow-stretched DNA interacting with condensins. They discovered that loops extruded by two condensins interact with each other, even when far apart. In close proximity, they observed that condensin complexes are able to traverse each other and form Z-loop structures. However, there have not been reported observations that validate the existence of such geometrical structures in vivo, because of the entanglement in the cell nucleus and the restriction in resolution of microscopy and filming. Nevertheless, our simulation is consistent with the observation in vitro, and provides strong support for the existence of Z-loops in cell nucleus.

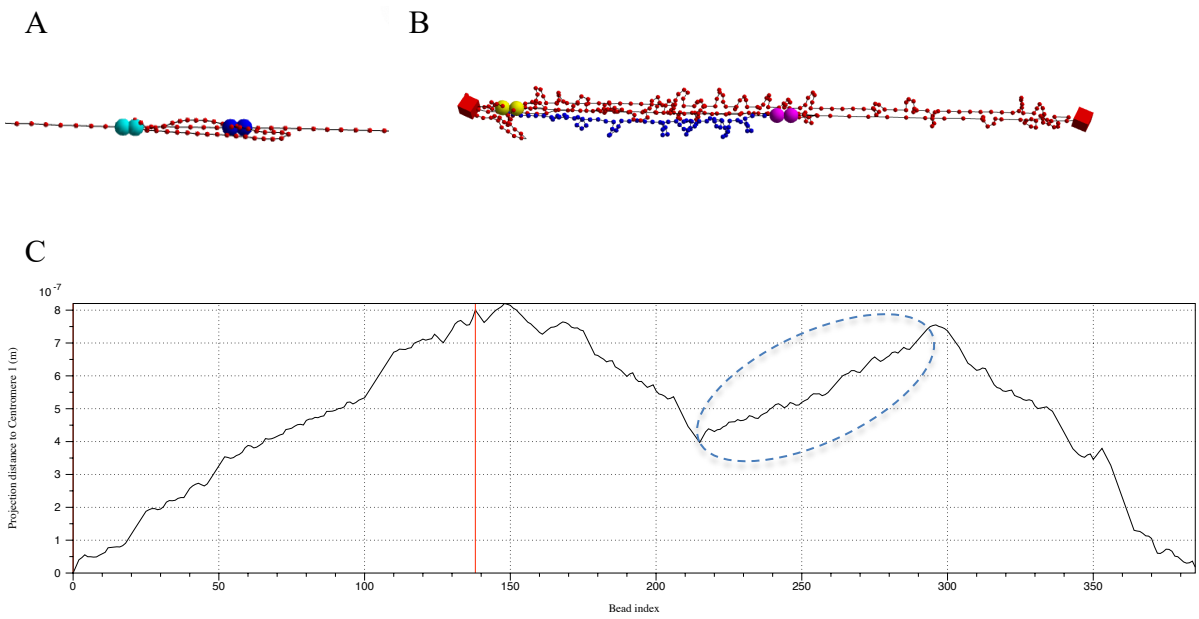



*Figure 3.10: Graphical demonstration of Z-loop structures*
(A) Graphical demonstration of Z-loop structure on a linear polymer chain formed by two interacting condensin loops. Large spheres with the same color code represent two attachment sites for a single condensin complex. (B) Snapshot of Z-loop structure in WT plasmid simulations. Blue labeled segment represents the backward segment caused by a Z-loop. (C) Projection distance from centromere 1 (Bead 0) as a function of bead index at the instance of Figure B. Two centromeres' positions are labeled as red vertical lines (Bead 0 and Bead 138). Circled region corresponds to inverse chromosome segment labeled in Figure B.

# CHAPTER 4: DYNAMIC CHROMOSOMAL CROSSLINKS CONDENSE AND STABILIZE STRUCTURAL HETEROGENEITY OF THE NUCLEOLUS


## Summary

The nucleolus is a large segment within Chromosome XII, itself the largest chromosome in budding yeast. The nucleolus is responsible for the synthesis of rDNA, yet its internal behavior has been challenging to resolve since it is comprised of regions of repeated DNA sequences. A chromosome model based on the statistical mechanics of polymers has been developed over the past quarter century, and utilized to explore the dynamics of chromosomes in budding yeast, and the interactions between chromosomes that are regulated by transient crosslinking and loop extrusion mechanisms due to structural maintenance of chromosome (SMC) proteins. In particular, transient crosslinking within the nucleolus by SMC proteins has been explored computationally, revealing phase separation of the nucleolus, clustering of genes with frequent exchanges between clusters, and compelling evidence, backed by experimental confirmation, that multiple rDNA loci on different chromosomes are indistinguishable from an intact nucleolus. In this work, a new method to identify the 3D numerical territory is developed to measure the volume of the nucleolus. Our simulations suggest that the nucleolus adopts the most compact structure if SMC crosslinks are weak and short-lived. Moreover, with the same high-frequency crosslink conditions, the multiple-loci situation has a similar volume as the single-locus case.


## Introduction

The eukaryotic nucleus has multiple sub-domains, each attaining distinct structure and functions. With the development of microscopy and chromosome engineering, the

biochemical composition of individual (membraneless) compartments has been greatly exposed, and the functions each compartment maintains through different cell stages has gradually been discovered. Among those compartments, the nucleolus obtains the largest volume. The existence of nucleolus structure was discovered early in the 1830s. Nucleoli are known as the site of ribosome biogenesis: synthesizing ribosomal DNA (rDNA) which was first noticed and studied in the 1960s. Nucleolar organization and dynamics can also be visualized nowadays, using fluorescent recovery after photobleaching (FRAP). However, very little is known about the mechanisms of how sub-structures are formed, and very few simulation approaches have been conducted to illustrate the dynamics in detail.

Polymer-based simulations of the nucleolus were inspired by biological observations achieved recently (Albert et al., 2013), where it was discovered that the dynamics of rDNA loci showed slower movement and larger differences with respect to the nuclear center, compared to the non-rDNA loci which performed consistent movement with the chromosome chains. This implied that the nucleolus affects the dynamics of chromosomes, and further investigation is demanded to uncover the underlying mechanisms. Wong et al. proposed a predictive computational model that simulates the dynamics of the yeast nucleus during interphase in 3D. In this model, the nucleolus was taken into consideration and was modeled by assigning 10 times larger size to the rDNA segments. This model of a heterogeneous polymer chain successfully explains a fair amount of experimental data (Wong et al., 2013).

The Bloom and Forest groups carried out numerical simulations of the dynamics of all chromosomes of a budding yeast cell during G1, confined within the nuclear membrane (Vasquez et al., 2016). Based on this model, a new approach was implemented that does not assume the nucleolus to be static (Hult et al., 2017). The nucleolus was modeled as a specific sequence on Chromosome XII that interacts with all chromosomes and enables transient crosslinks between close rDNA segments, in order to represent interactions with SMC

proteins or HOM1, a high mobility class of protein participating in rDNA regulation in the nucleolus.

Our analysis was based on the crosslink model. We seek a method to transform simulation results to 3D signals. The validity of the method can be tested if we take slices from the 3D signal to obtain simulated 2D signals. The simulated 2D signal images can be directly compared to biological observations from microscopy. The volume of the nucleolus is represented by the volume of the 3D signal. This method grants researchers a direct observation to the 3D volume of the simulated dynamic nucleolus, and becomes the baseline for further metrics and studies.

## Materials and Methods

### *Chromosome Modeling Approach*

A classical, Rouse-like, bead-spring polymer model is applied to simulate all 16 chromosomes within the nucleus of live budding yeast cells. All chromosomes are tethered at measured locations in their interior to the common centromere that is tethered to the spindle pole body, with 32 chromosome arms tethered to a total of 6 telomeres on the nuclear wall. The chromosomes are discretized to form the total bead-spring polymer system. Each bead represents 5 kbp of DNA, and adjacent beads are connected by nonlinear springs. The length of each polymer chain is measured in vivo (Hult et al., 2017). Moreover, all chromosomes are confined in a sphere of radius 1 $\mu m$, which represents the nuclear membrane. The geometry follows the Rabl configuration (Therizolsa et al., 2010).

The motion of each bead *i* is governed by the following equation in 3D:

$$\zeta \frac{d\boldsymbol{x}^{(i)}}{dt} = \boldsymbol{F}^{(i)}_{\boldsymbol{spring}} + \boldsymbol{F}^{(i)}_{\boldsymbol{excluded\ volume}} + \boldsymbol{F}^{(i)}_{\boldsymbol{thermal}} + \boldsymbol{F}^{(i)}_{\boldsymbol{confinement}}, \tag{4.1}$$

where $\zeta$ represents the global fluid drag coefficient following Stokes law. This system is dominated by drag, and hence a non-inertial approximate model is applied, following the so-

called free draining approximation. $\boldsymbol{F_{spring}}$ represents the spring force following a worm-like chain force law (John F. Marko & Siggia, 1994) to resemble elasticity of DNA. $\boldsymbol{F_{excluded\ volume}}$ and $\boldsymbol{F_{confinement}}$ are restricting forces that prevent the target bead from colliding with other beads and prevent the bead from escaping the nucleus, respectively. Whenever the bead comes within a prescribed distance of another bead or the membrane, a repelling force is applied on the bead. Last but not least, the system is subject to entropy fluctuation, which is simulated by applying a random white noise force to each bead. For details of the forces and choice of parameters, refer to (Vasquez et al., 2016).

The simulation program was coded in C++ using *DataTank* as graphical interface. Customized parameters and geometrical initializations are predefined in *DataTank* script before running the simulation. *DataTank* allows real-time visualization to locally run simulations. Output files could either be imported to *DataTank* and analyzed using its embedded analytical modules, or be exported as open source data files and further analyzed by using alternative methods.

*Nucleolus*

The nucleolus is represented in the model by a specified location of 361 beads along arm 2 of Chromosome XII, which represents published results of roughly 1.8 million bp in length. Another specification is given for a split nucleolus configuration: the nucleolus beads are evenly divided into two portions, half on arm 2 of Chromosome XII, half on arm 2 of Chromosome III. The exact positions for each configuration are consistent with experiments of the Bloom lab.

In terms of mechanistic properties, nucleolus beads are the same as beads on other chromosomes. All beads, springs and force laws are uniform. The only difference is that nucleolus beads are preferential attachment sites for crosslinks, in some simulations exclusive

binding in the nucleolus, in other simulations some fraction of non-nucleolus beads are active in forming crosslinks. The crosslinks transiently bind to and unbind from chromosome beads; while bound, the crosslinks are modeled as dynamic springs between attachment sites. These crosslink springs obey the same worm-like chain spring law, yet are significantly stronger than the springs between neighboring beads on the chromosome spine. Crosslinks form stochastically, allowed if two candidate beads are close enough. The probability of formation, and the on/off duration for each formed crosslink, are essential parameters that greatly influence the simulation results.

The validity of the base model and the nucleolus model was statistically tested against experimental results. Methods include MSD analysis of beads among others; for detailed comparison results, refer to (Hult et al., 2017; Vasquez et al., 2016).

*Signal Transformation*

The simulation outputs a time series of 3D coordinates for all the beads. Our focus lies on the nucleolus beads, which are labeled consistent with GFP labeling in experiments. Unlike experimental results where the signals are blurry due to the restriction of resolution, the simulations output the exact positions of the nucleolus beads. Therefore, a signal transformation tool is necessary to convert the simulation data to experimental observation images.

A *DataTank* pipeline is created to convert simulation output to a 3D point-spread function by applying a point-spread function to each bead to generate a 3D signal. The choice of the point-spread function is consistent with the microscopic images created in the Bloom Lab to guarantee that minimum bias is introduced. This process generates a global intensity map discretized at each spatial unit in 3D. Dark regions acquire 0 as the mapped values and large values correspond to bright pixels.

After we get the global intensity map, we need a method to determine the precise contour of the signal. A number of image thresholding methods have been tried; the improved median-based Otsu thresholding method (Yang et al., 2012) is applied to acquire the appropriate 3D signal. This method identifies the signal contour by learning the cumulative distribution of the intensity and makes decisions based on statistics like the median. Compared with traditional Otsu thresholding method where the primary working statistics is the mean, the modified method benefits from its robustness. This is essential for our particular case, since the distribution is often heavy tailed. Sometimes multiple rounds of the thresholding method are applied, to avoid error from multipolar distributions. A universal criterion would be ideal, but is unlikely to be applicable. The cause is that the differences in geometry and distribution of beads will lead to extraordinary variation in signal intensity after mapping. This dilemma is also encountered in processing experimental images.

The 3D simulated signal is the ultimate result we seek for. For further biological studies, one can take the max projection of the 3D signal to generate a 2D planar signal. By taking appropriate image processing procedures, the 2D planar signals generated by numerical simulations can be directly compared to the biological observation images in microscopy. For concise description on the manual, see Appendix 1.

## Results

### *Yeast nucleus simulation results*

The simulation initializes with linear configuration for all chromosomes (Figure 4.1 A). Since all 16 pairs of chromosomes share one common centromere and 6 telomeres, some chromosomes obtain identical initialization and therefore are not distinguishable in the starting stage. The nucleoli portion are discretized into beads on Arm 2, Chromosome XII labeled in green, which simulates the CDC14-GFP fluorescent labeling of nucleoli. As the

simulation progresses, the thermal fluctuation breaks down the conformation. The chromosomes spread across the whole nucleus in a short period of time due to intense diffusion.

Based on the knowledge of Rouse chain models, random coils will form for pure Rouse chains that freely diffuse. However, instead of an equilibrium of pure random coils, the chromosomes have a tendency of localization. After reaching equilibrium, the nucleolus beads spread in a local region near its initial position. A natural explanation lands on the additional laws that govern this system: the geometrical confinement and the excluded volume.

The geometrical confinement consists of two parts: the tethered ends and the bounding membrane. Both are reasonable sources for localization. External forces propagated from the thethering sites will be applied to the chain to prevent it from escaping its local region. Furthermore, the excluded volume mechanism allows one chromosome to interact with the other chromosomes. In such an overcrowded space, individual chromosomes require tremendous energy to squeeze through the cloud of particles and reach a farther space. However, the thermal dynamics, as the only source of external energy, is not sufficient for such a behavior. Both factors restrict the mobility of chromosomes, ending up with localization and formation of sub-domains.

However, through chromosomal engineering, scientists have generated rDNA-split chromosomes (Mikus & Petes, 1982) and surprisingly the split nucleoli are not distinguishable from WT single-nucleolus nucleus. This argues against the localization properties for chromosomes in the nuclear environment. Presumably, if two strands of rDNA locate separately, they will fluctuate in their distinct territories, forming two clusters of signals in isolation. But only one cluster is observed in vivo. Remarkably enough, the signals are in comparable size for the two situations. This leads to further interest in exploring the

mechanisms of the nucleolus, and provides motivation to simulate the unique dynamics and interactions of the nucleolus.

We simulate the interactions within the nucleolus by imposing transient crosslinks between rDNA segments. These crosslinks dynamically form between two close beads, and maintain the strong binding mechanism for a predefined period of time, $t_{on}$. After staying attached, the crosslinks deactivate and remain muted for a different time period, $t_{off}$. Notice that when an rDNA segment crosslinks with another segment on the same chromosome, a loop will form. The crosslinks are only allowed to form in the nucleolus. Then we tune the thermal fluctuations together with crosslink timescales to observe the dynamical impact these crosslinks bring to the system, and try to uncover the mechanistic explanation behind the split nucleolus situation.

Surprisingly, we observe a huge impact on the localization of the nucleolus region when crosslinks are imposed. Moreover, we observe clear community-organization behavior for nucleolus beads when the transient crosslinks switch phases rapidly (Figure 4.2). In simulations, we tune the crosslink probability to bind and once bound, to unbind, by varying $t_{on}$, and $t_{off}$ is set to $1/9$ of $t_{on}$. We observe a gradual merging effect for split nucleolus situations as the persistence times of crosslinks decreases. For extremely long-lived crosslinks ($t_{on} = 90s$), we observe obvious localization among the nucleolus beads, compared with no crosslink simulations. The beads are more confined in a nearby area and the occupying volume is reduced. However, in the split nucleolus case, no significant merger was found between the two split regions of rDNA. There are occasional crosslinks happening between beads in distinct regions, but the overall distributions are still separate in space.

As the crosslink duration drops, more intense interactions occur among nucleolus beads. Enormous numbers of local interactions propagate to a global structural phase change, and makes the nucleolus more compact. In split nucleolus cases, the split regions start to

merge into one crossing the energy barrier (Figure 4.2). In $t_{on} = 9s$, split nucleolus conditions, 4 out of 10 simulations end up converging to a single-nucleolus equilibrium. In more short-lived crosslink simulations, nearly all split nucleolus simulations evolve into a single nucleolus equilibrium. Clearly, the active crosslinks contribute sufficient energy for this merger.

In rapid forming, short-lived crosslink situations where $t_{on} < 0.9s$, we start to observe community behavior. A small number, usually 10-25 beads, gather and form a dense cluster. These clusters persist on timescales sensitive to the value of $t_{on}$, and vary from different runs. This observation is non-intuitive: we introduce pairwise interactions that yet result in community-scale (many beads) organization. This remarkable observation is further studied and characterized through network-based analyses (Walker et al., 2019).

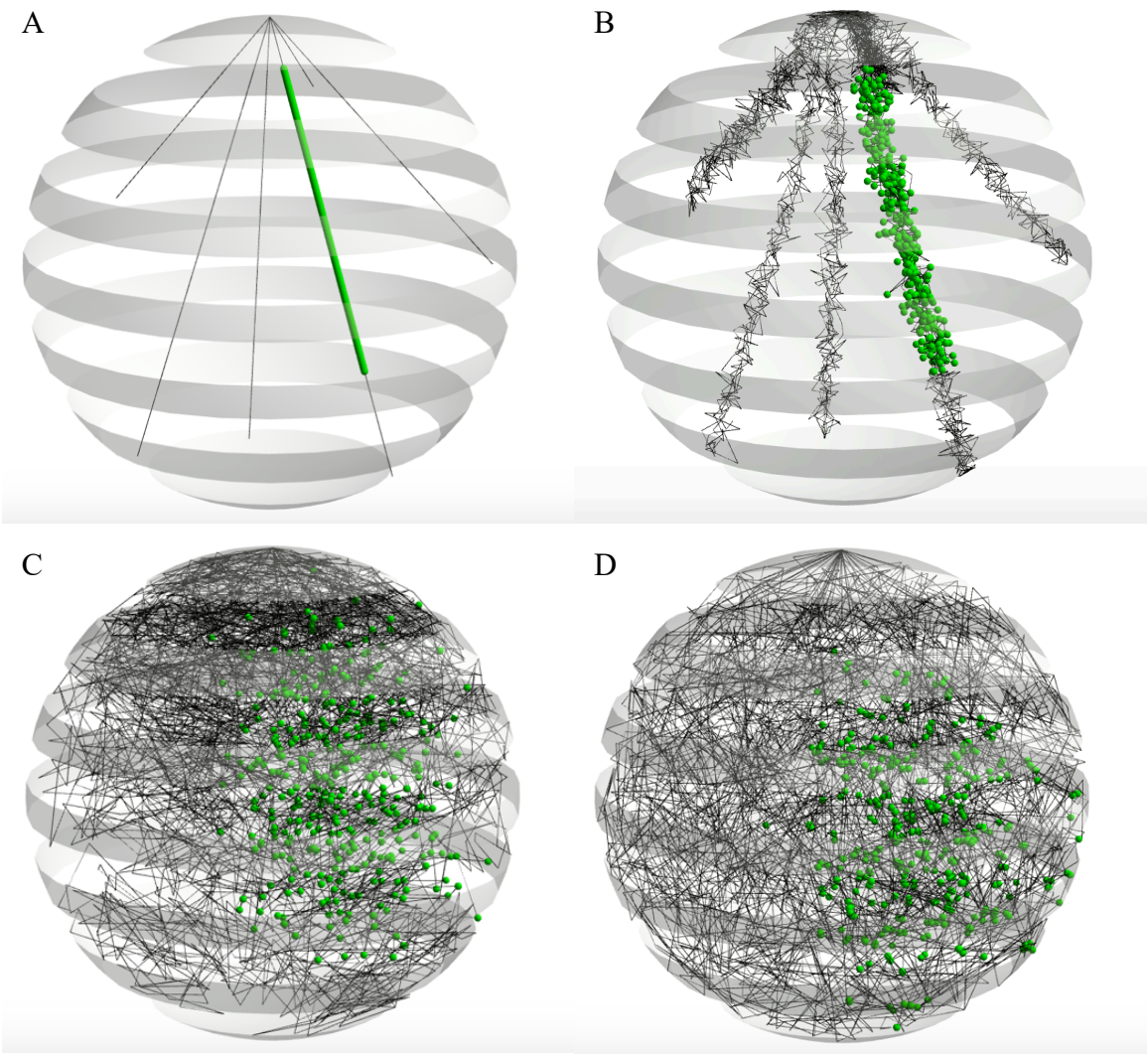
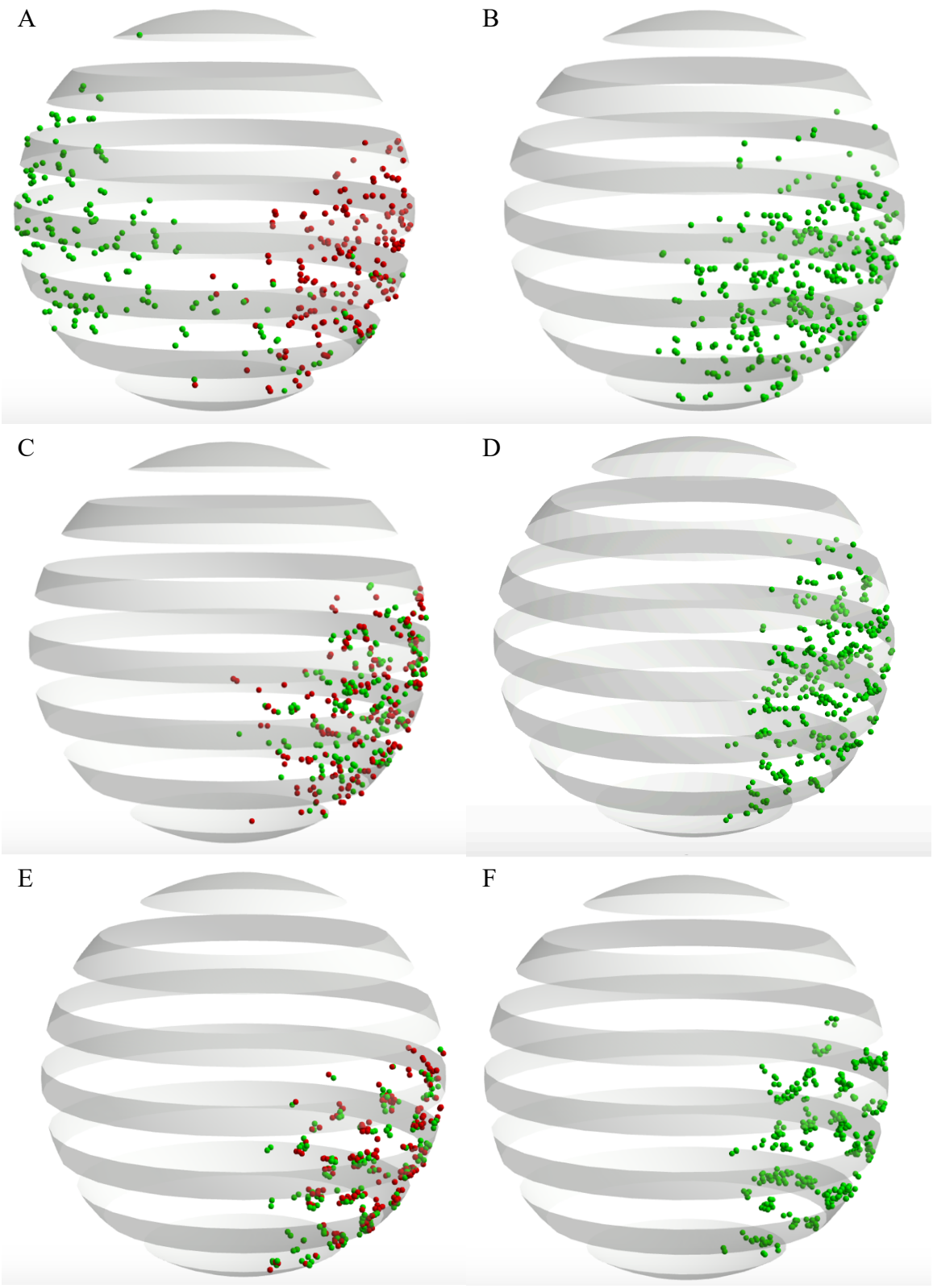



*Figure 4.1: Simulated chromosomes in yeast nucleus*
Simulated yeast chromosome at (A) $t = 0s$ (B) $t = 0.2s$ (C) $t = 10s$ (D) $t = 1000s$. All 32 chromosomes are included in this simulation. Green beads represent discretized chromosome in nucleoli region for single nucleolus condition. The large gray sphere is the virtual nuclear membrane. Crosslinks are not imposed.


A
B
C
D
E
F

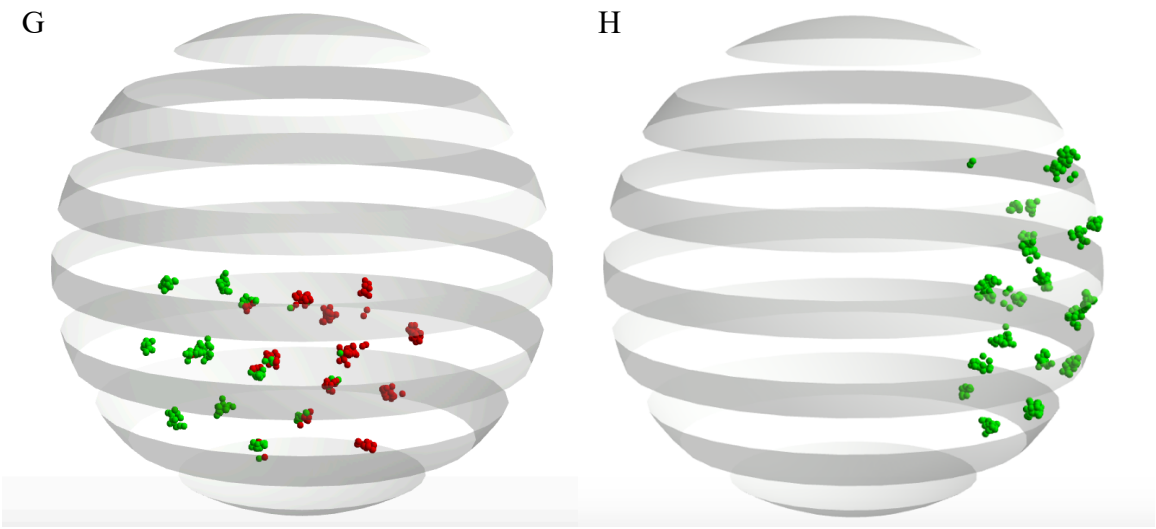



*Figure 4.2: Snapshots of simulated nucleolus beads*
Simulation snapshots of nucleolus bead spatial distributions against varying crosslink frequency at $T = 3700s$. (A) (B) $t_{on} = 90s$. (C) (D) $t_{on} = 0.9s$. (E) (F) $t_{on} = 0.17s$. (G) (H) $t_{on} = 0.09s$. Left column (Figure A, C, E, G) shows split nucleolus situations. Right column (Figure B, D, F, H) shows single nucleolus situations. For split nucleolus situations, discretized beads from distinct portions of the nucleolus are differently color-coded. Individual simulations are independent of one another.

*Simulated signal processing results*

To develop a methodology to generate simulated signals comparable to experimental results, we apply a 3D point spread function to each bead in the simulation. The function values are superposed at locations near multiple beads, forming visible signals. Isolated beads emit signals with less intensity. This process generates a global intensity map discretized at each spatial unit in 3D. Dark regions acquire 0 as the mapped value and large values corresponds to bright pixels.

After we get the global intensity map, we apply image processing methodologies to determine the contour of the signal. The result, a 3D bulb of signal, is the signal emitted by the labeled beads (Figure 4.3). 2D simulated images can be obtained by specific projection methods (often max projection) from the 3D signal onto a 2D plane (Figure 4.4). Direct comparison can then be conducted for further studies.

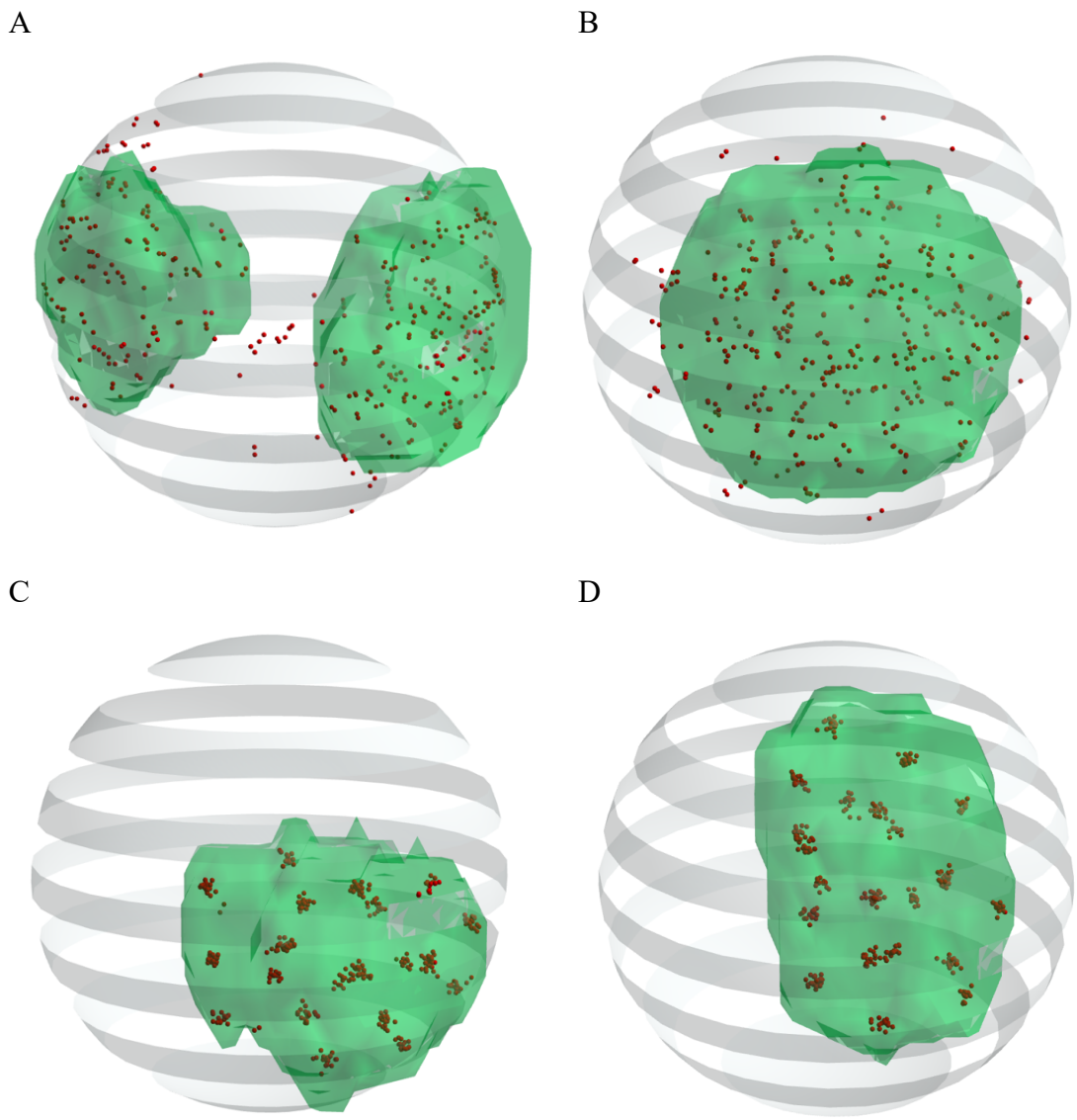



*Figure 4.3: Simulated 3D signals of nucleolus*
Simulated 3D signal from nucleolus at $T = 3700s$. (A) Split nucleolus, $t_{on} = 90s$. (B) Single nucleolus, $t_{on} = 90s$. (C) Split nucleolus, $t_{on} = 0.09s$. (D) Single nucleolus, $t_{on} = 0.09s$.

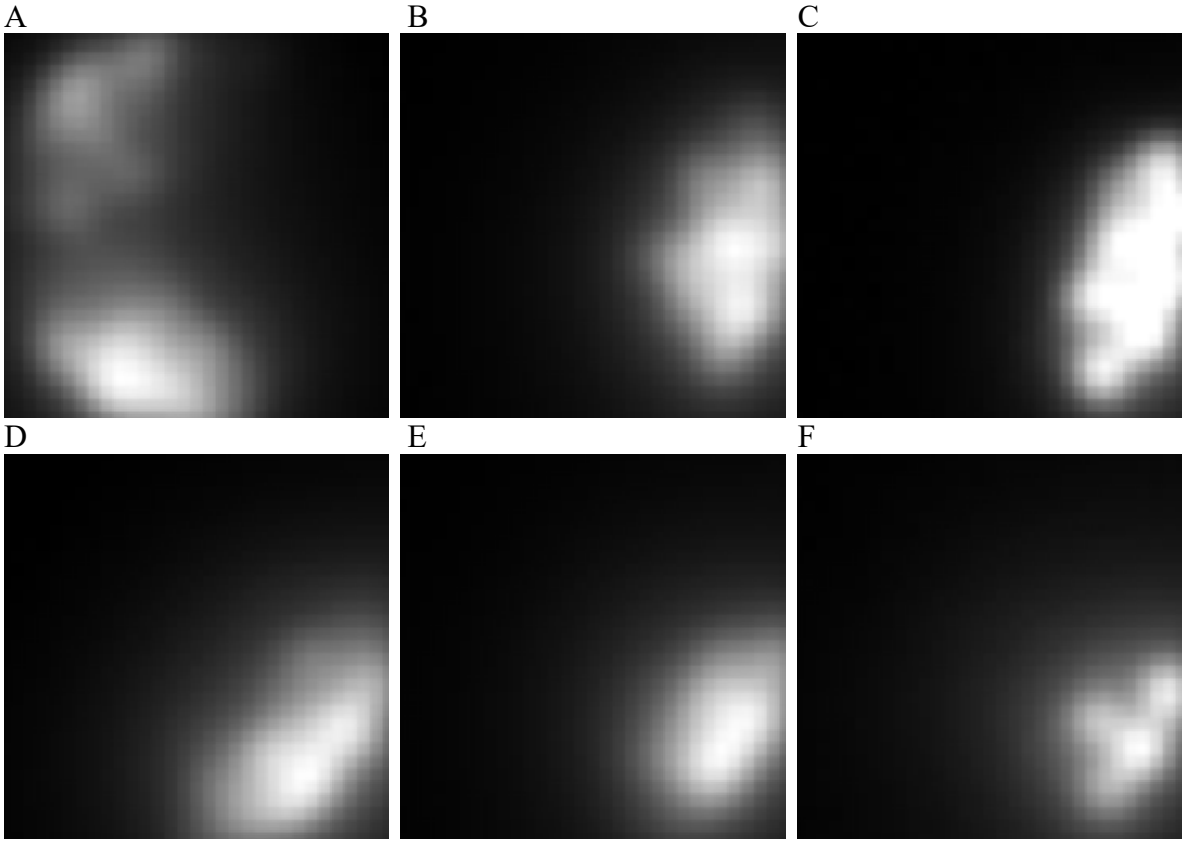



*Figure 4.4: Simulated 2D signals of nucleolus*
Simulated 2D signal from nucleolus at $T = 3700s$. (A) (D) $t_{on} = 90s$. (B) (E) $t_{on} = 0.9s$. (C) (F) $t_{on} = 0.09s$. Upper row (Figure A, B, C) are in split nucleolus condition. Bottom row (Figure D, E, F) are in single nucleolus condition.

*Community structure compacts the nucleolus.*

We measure the nucleolus volume by measuring the volume of the 3D signal region. Although bias is unavoidable, this approach will provide reasonable indication on the true nucleolus volume.

We first compare the volume of the nucleolus across a range of crosslink timescales (Figure 4.5 A). A significant increase is shown in the plot at labeled frequency $r_0$ where $t_{on}$ is roughly $1s$. An average volume of $0.55\ \mu m^3$ grows to $0.7\mu m^3$ for both split and single cases when the crosslink frequency drops below $r_0$. This coincides with the observational fact that community behavior starts at $r_0$, and thus confirms the hypothesis that community organization induces a compact, stable nucleolar area. Additionally, from subtle observations on the simulation configuration, we observe significant localization when clusters appear. Instead of intense fluctuating behavior for beads on a common Rouse chain, beads within one cluster tend to oscillate within the extent of the community or switching between nearby communities. This leads to a more stable global configuration. If we focus on the explored area of the nucleolus, the influence of the community behavior will be more significant.

Without the phase change in community behavior, varying the crosslink frequency does not bring essential change to the volume of nucleolus. No global tendency is discovered in low crosslink frequency simulations where no clustering behavior is detected. However, the compactness of the nucleolus scale inversely with the frequency, which can be characterized by the cumulative volume the nucleolar beads have explored in time. With slow, long-lived crosslinks, the clusters melt down and very little sub-structures arise, and the geometrical confinement of individual bead will drop, leading to a more widely-spread nucleolus. However, it is worth mentioning that the nucleolar volume does not indicate the dynamics of the nucleolus. Rapid, short-lived crosslinks induce clustering and a compact geometry, but also reveal an optimum timescale for the mixing of all clusters where beads

swap between communities. Under such a timescale, the nucleolus occupiexws less volume in space but possesses intensive internal dynamics (high frequency crosslinks).

*Split nucleolus does not lead to significant volume change in nucleolus*

In Figure 4.5 B we do not see evidence of variation in nucleolar volume between split nucleolus and single nucleolus, independent of whether there is heavy clustering behavior. To dive deep into the analyses, we constructed 20 full simulations at crosslink frequency $t_{on} = 9s$. Half of them obtain a single nucleolus region on Chromosome XII. The other half are initialized with two separate nucleoli on two separate chromosomes. The reason for choosing specific $t_{on}$ is that both split nucleolus equilibrium and single nucleolus equilibrium are possible states under this condition. In the simulation output, 6 out of 10 runs in split nucleolus conditions finish with a single nucleolar region. The split nucleoli remain separate throughout the other 4 runs. Therefore, we get three classes of results: the single nucleolus case as the control group, the "split single" group where the split nucleolus merges into one at equilibrium, and the "split split" group where the split nucleoli do not merge.

Across all groups, the mean volume is comparable at around $0.7\ \mu m^3$. The "split single" group obtains a slightly higher mean of $0.72\ \mu m^3$, but the large variance denies a statistical distinction between cases. The "split split" group attains a significantly higher variance, which is caused by occasional changes in conformation aroused by crosslinks across separate regions. This suggests the difference between split nucleolus condition and single nucleolus condition cannot be distinguished by nucleolar volume. An image of a single nucleolus is indistinguishable if it has a single nucleolus or separate nucleoli. This confirms with biological observations in cells with rDNA at ectopic chromosomal sites (Oakes et al., 2006).

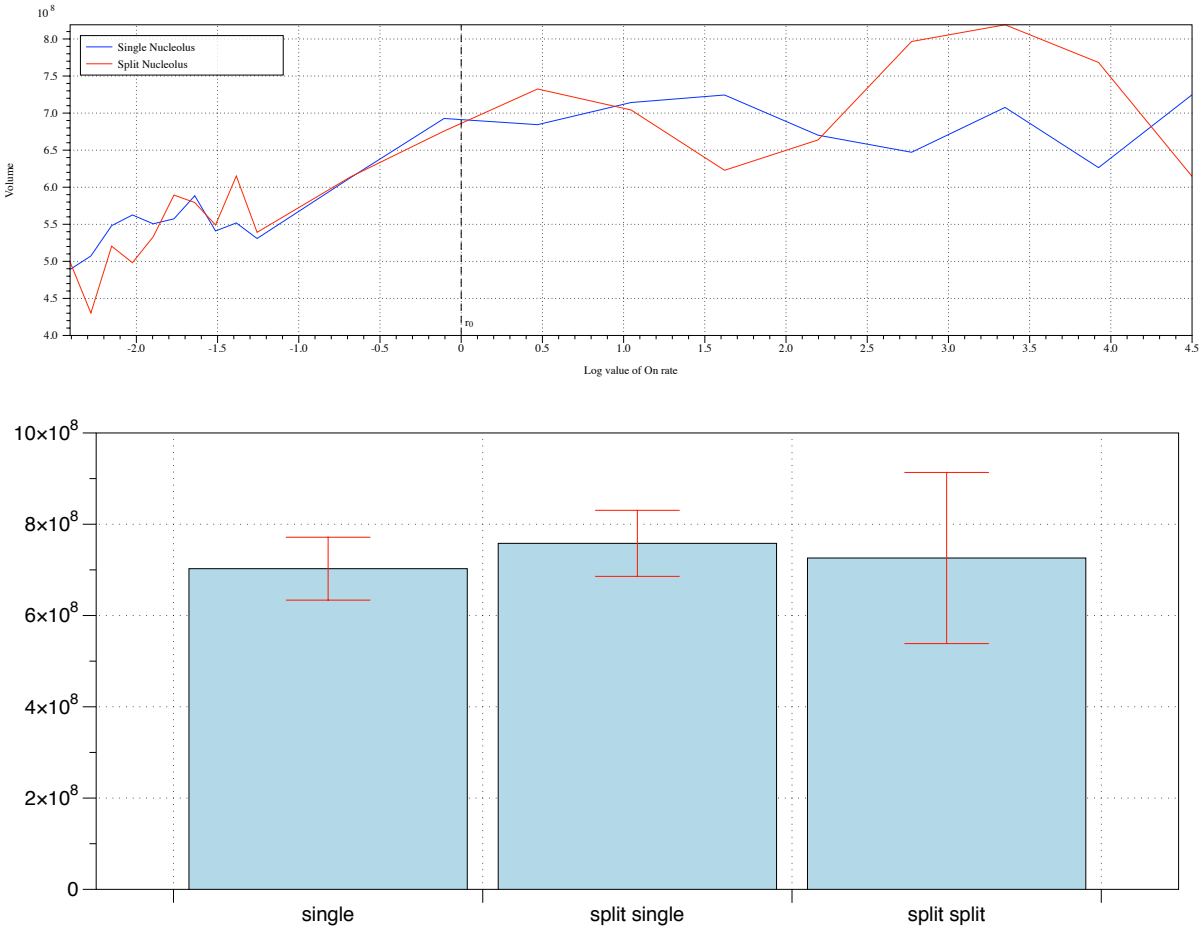



*Figure 4.5: Nucleolar signal volume comparison plot*
(A) Simulated signal averaged volume plot against $t_{on}$ for both single nucleolus case and split nucleolus case. The mean volume was calculated from multiple samples drawn from one run. (B) Simulated signal averaged volume bar plot against nucleolar condition. Red error bar refers to standard deviation. 10 runs in single nucleolus condition and in split nucleolus condition are sampled and analyzed, respectively. In split nucleolus conditions, 6 runs reach a single nucleolus equilibrium. The mean and SD for "split single" bar are sampled from these runs. The data for "split split" bar is sampled from the remaining 6 runs that end up without merger. The volume units are $nm^3$ for both plots.

**Discussions**

Previous studies in the Forest group follow the path of revealing the structure and dynamics of yeast chromosomes. During interphase, entropy and other global interactions dominate the formation of territories (Vasquez et al., 2016). Based on this, we develop simulations for specific interactions of different compartments. The simulation for nucleolar association by transient crosslinks extends the study (Hult et al., 2017). Focusing on the community behavior discovered in simulation, a multilayer modularity optimization scheme is applied to detect clustering and mixing between clusters (Walker et al., 2019).

The simulations of the nucleolus offer a candidate model that explains biological observations. Under this hypothesis, community behavior was discovered to support biological intuition that sub-domain clusters provide compactness and overall stability of the nucleolus. The formation of compartments favors observations of nucleolus regions in the nucleus (Boisvert et al., 2007). The crosslinking behavior is introduced by structural proteins. Many studies have also addressed the role proteins play in phase separation (Hancock, 2012). However, we have not specified particular proteins or other structural molecules that are candidate carriers for the transient crosslinks. We look forward to biological breakthroughs to unveil mechanisms for structural proteins or RNA in this scale.

Being one of the hypothetical local interactions in the genome, transient crosslinks give rise to formation of sub-domain structures. Furthermore, there are extra interactions to be explored and embedded into our simulation. One recent mechanism is loop formation. In our model, loops are mediated by transient crosslinks, being static and only occuring in specific regions. Looping is considered a common phenomenon across the genome, and multiple hypotheses predict loops across various scales. We learned and modeled loop extrusion behavior caused by condensins, but these small-scale proteins are shown capable of forming up to 10 kbp of loops. In other works (Bohn & Heermann, 2010), formation of loops

is solely accomplished on the basis of diffusional motion. Therefore, alternative models can be developed to impose various structural changes in chromosomal organization.

The assumption of a preferential crosslinking mechanism elevates the stability of the chromosomal structure in phase separation. The study is versatile and potentially can be modified and applied to similar structures. For example, there are other regions in the genome with fast exchange of proteins, such as Cajal (Handwerger et al., 2003).

# CHAPTER 5: FUTURE DIRECTIONS

## Completion and Extensions of Simulations on Chromatin Dynamics

The conditionally dicentric plasmid chromosome which Bloom's lab used is proven to be a useful tool in studying how DNA-binding protein complexes impact chromosomal structures and dynamics. Its isolation convenience observations and mutations in experiments, and its reasonable geometrical scale acts as a good carrier to reflect the protein interactions. Based on this, numerical simulations play a key role to make full use of this biological tool. Appropriate simulations not only reveal chromatin dynamics at a fine scale with low cost, but also offer intuitions and insights back to biology. There are a number of ways the simulation pipeline can be improved, in the aspects of either credibility or performance.

The dynamics of DNA-binding protein complexes is essential yet not taken into consideration in our model. However, the dynamics of nucleosomes formed by histones exhibits dominating influence on the dynamics of DNA. Nucleosomes formed by core histones can be further arranged to form high order structure (Johansen & Johansen, 2006; Stephens et al., 2011). A chain of nucleosomes can be arranged into a compacted fiber with extensive compaction depending on H1 histones (Luger et al., 1997). Besides, the positioning and interactions of nucleosome are also heterogeneous (Teif & Clarkson, 2019). These studies showed the observational results, but clear mechanistical demonstration is still at demand. This suggests the high priority of including the dynamics of protein complexes, in order to learn the causation of higher-order structures of chromatin and the impact of the organizations on DNA.

To embed full hydrodynamical environment is a reasonable idea to achieve better credibility. Since cell nucleus is filled with nucleosol, a gel-like fluid substance similar to cytosol, and crowded with full genome, intense protein-DNA interactions and DNA-DNA interactions occur and are propagated through fluids. Embedding hydrodynamical environment will provide the platform for these interactions and influences, and will provide insights on the dominancy of hydrodynamics in small-scale systems. Immerse boundary method (Peskin, 2002) is a potential methodology that examines the dynamics of fluid influenced by boundaries. By iteratively update the dynamics of solid objects (in our case the chromatin) and the fluid solvent, a hydrodynamics-embedded chromosome dynamics simulation can be achieved. Presumably, the complicated solid boundaries will cause numerical inefficiency. Thus, the challenge is to optimize and develop an efficient simulation program for simulating cellular systems. Nevertheless, it is a debatable proposal to eventually include hydrodynamics into the model. Whether hydrodynamical forces would be a significant component enough to cause impacts in the system remains unknown, and requires analyses or trial simulations. The same argument goes to additional environmental factors, such as electrostatics.

Furthermore, with the development of biological observations and studies on cell nuclei, additional needs for simulation may become necessary, either alternative types of DNA-binding proteins, or models of existing proteins with distinct properties. For instance, spindle pole bodies are observed to be oscillating instead of staying still, and this oscillating behavior is hypothesized to be depending on tension. Presumably, this tension-related behavior will influence the behavior of tension-sensing proteins. To embed this idea into simulation requires direct modification on the simulation geometry. Predictably, this modification requires the ability of the simulation to be well organized and easily modified. Fortunately, the simulation program is developed in C++ and is well object-oriented

designed, which greatly benefits modulization and modification. Alternative modification occurs for simulating loop extrusion activity of condensins. The alternative hypothesis states that a condensin complex extrude loops sequentially along the chromatin. This deviates from our model, where non-sequential extrusion is accepted and is even preferred to explain sudden disappearance of loops.

## APPENDIX 1: CHROMOSOME SIMULATION MANUAL[1]

The simulation pipeline uses *DataTank*, a numerical working environment, as the user interface. *DataTank* is developed by David Adalsteinsson; researchers can go to the following website for a user license and to download the software https://www.visualdatatools.com/DataTank/. *DataTank* provides powerful visualization and analysis tools that are compatible with various types of data. *DataTank* employs embedded data-oriented functions and is itself a visual programming environment. Programs created by standard utilities or the built-in C++ framework can be launched in *DataTank* for numerical purposes. The simulation results are parsed by *DataTank* in real time for downstream data visualization and analyses.

1. Plasmid_Chromosome_Simulator.tank: The main file for plasmid chromosome simulation pipeline, which contains: 1) The interface to enter input parameters and settings. 2) The port connecting to the simulation program. 3) The C++ script attached to the .tank file. 4) The visualization module.
2. Data_Analysis.tank: The file reads the output file from Plasmid_Chromosome_Simulator.tank and contains various high-performance data analytics tools that analyze the output. Methods include but not limited to: MSD analysis, Radius of Gyration analysis, persistence length analysis, distribution analysis, contact map analysis and more.
3. Yeast_Chromosome_Dynamics_Simulator.tank: The main file for the simulation pipeline contains: 1) The interface to enter input parameters and settings. 2) The port

[1] This chapter previously appeared in SMC Complexes. The original citation is as follows: Lawrimore J, He Y, Forest G M, et al. Three-dimensional thermodynamic simulation of condensin as a DNA-based translocase[M]//SMC Complexes. Humana, New York, NY, 2019: 291-318.

connecting to the simulation program. 3) The C++ script attached to the .tank file. 4) The visualization module. 5) Modules for stochastic analyses.

4. Microscope_Simulator.tank: The microscope simulator pipeline functions to convert the numerical results to experimental microscope images. This enables direct statistical comparison of geometries between simulated chromosomes and experimental observations.

All DataTank scripts and attachments are available at https://github.com/heyunyan19930924/Yeast-Chromosome-Dynamic-Simulaion.

**Run Simulation Programs**

We now show how to run the simulation program using DataTank. After the simulation data is generated, we show ways to store the results in different file formats. We show how to visualize the results as the simulation is progressing. Lastly, we introduce scripts for data analytics of the simulated, all chromosome, time series, consisting of a Microscope Simulator script that passes the simulated data through a microscope imaging pipeline, and a Community Detection script that detects dynamic gene clusters (communities) at the scale they exist.

Both plasmid chromosome simulator and yeast chromosome simulator share the same pipeline for data acquisition and data visualization. Therefore, the manual is shared. Here we only precisely display the manual for yeast chromosome simulator. For plasmid chromosome simulator, only unique features will be introduced.

*Yeast Chromosome Simulator*

1. Download and install *DataTank*. Acquire the license by following instructions on the website. Download the simulation script Yeast_Chromosome_Dynamics_Simulator.tank.

2. Run Yeast_Chromosome_Dynamics_Simulator.tank through *DataTank*. Set inputs before running the simulation. Default setting simulates DNA dynamics inside the nucleus of a budding yeast cell. First set the initial geometry by inputting the number of arms, lengths of each arm, and connections at the nuclear wall in the "lengths" and "connections" modules. By default there are 32 polymer arms simulating 16 chromosomes. "Coefficients" dictionary contains parameters regarding to dynamics, including simulation scale, persistence length of polymers, environmental parameters, and so on. "Evolution" dictionary controls the simulation time scale, time step, and which timesteps to save (the save time "stride"). The fundamental model introduced so far already builds the basis for polymer dynamics simulation, and there are additional features specifically for simulating chromosomes. Beads are subject to a repulsive potential and the corresponding parameters are in the "Excluded volume" dictionary. The nuclear membrane is modeled as a spherical surface and all arms are restricted within. The membrane location and volume are set in "Void formation" dictionary. By default, both ends of all polymer arms are attached to fixed sites on the membrane. Moreover, this work simulates chromatin interactions as transient crosslinks in the nucleolus, and the nucleolus is modeled as a specified region (the default is on Chromosome XII) specified in the "Terms" dictionary. Users can toggle all modules listed above to modify the simulation parameters.
3. The simulation can be run in two different ways. The user can either run the simulation locally or submit it to a remote computing cluster.
    a. Local. This approach utilizes the local machine's CPUs, and therefore potentially prevents users from multitasking, especially for long simulations, but it allows users to view the results simultaneously. One can create a "Task" module, add dependencies to all relevant dictionaries simply by dragging the

dictionaries in the module. By linking the C++ simulation script to the module as the kernel, the simulation is ready to run.

b. Remote. Submit the simulation and run it on a server with legal access to avoid local occupancy. This approach forbids simultaneous observation; however, one can download the results and visualize them locally at any time, even when the job is still running. To create a job, DataTank offers the "Parameter run" module that does the submission and online manipulation. Likewise, once the "Task" module is created, one can link all dependent parameters and simulation script to the task. One can conduct multiple runs in each submission by setting variables as "Time sequences". DataTank automatically recognizes the format and creates multiple runs for all time values.

4. *DataTank* is only compatible with Xcode in MacOS as its C++ program compiler. It also provides built-in DTSource library as data structures and ports specifically for *DataTank* modules. The simulation C++ script is developed based on the DTSource library and is attached to Yeast_Chromosome_Dynamics_Simulator.tank.

   a. Set the application to the attached script in "Parameter run" or "Task" module by clicking the "Set" button; find the script under Yeast_Chromosome_Dynamics_Simulator.tank file.

   b. Compile or profile the C++ script in Xcode, and one will find that in the scroll down list "Support", either the "Debug version" or the "Release version" has been lit up.

   c. Select either the "Debug version" or the "Release version" as the project.

   d. If the simulation is being run locally, click the "Run" button and the simulation starts to progress.

e. If the simulation is being submitted to a remote computing cluster, click the "Create" button to create a Parameter_run.dtask file that stores all required scripts and commands for submitting the job. Navigate to "Task > Run remotely > Add machine" and enter server information and access. Now click the "submit" button, and the job is successfully submitted to the server. The output data is also stored in the same Parameter_run.dtask file. While the file is open, DataTank periodically checks with the server for progress and downloads temporary results.

5. To get the simulation results, the approach depends whether the simulation is run locally or online.
    a. For local simulations, the output is saved by DataTank into a local cache. By accessing the "Task" module itself, we parse the results. If one intends to store the output locally, make a switch in the "Directory" line from "Temp" to a custom directory in the "Task" module before running the program. The output will be stored in a custom .mat file as the simulation progresses.
    b. For online simulations, the output data is saved inside Parameter_run.dtask file.

*Plasmid Chromosome Simulator*

1. The geometrical initialization of plasmid chromosome simulator is less complex than yeast chromosome simulator. The total number of beads included in the simulation is tunable. However, there are a large number of parameters regarding to environmental settings and modules to be included.
2. The functional proteins, namely the histones and condensins, possess their own parametric dictionary. All properties corresponding to the structural proteins are

specified under “histone” and “condensin” dictionary, respectively. For example, the extrusion speed is tunable via “condensin> judgeTimeSeconds”. To turn the structural protein modules on and off, go to “flags > histone/condensin” and change the flags accordingly.

3. A unique functionality in plasmid chromosome simulator is that it allows heterogeneous polymer chain. Navigate to “differentSpringProperty” dictionary, and it enables a modification of stiffness within a portion of beads on one chromosome. For instance, half of the chromosome maintains a persistence length of 50 nm by default, while the other half obtains a high persistence length of 500 nm. This is useful to learn the impact polymer stiffness will cause on condensin distribution.

**Access Simulation Results**

1. Simulation data stored as .mat file contains the coordinates of each bead at each timestep in a table and stores all tables in a corresponding .mat file. *DataTank* can parse the .mat file as input and convert to a time series of “3D Path”, a built-in data structure in the DTSource library. By creating a “Data File” module in *DataTank*, the .mat data can be read in the module. Three variables are contained in the .mat files, and the position information are saved as variable “Var”.
2. Simulation results saved in .dtask files can be read by “File Name” module in *DataTank*. Choose “From Parameter Run” from the scroll down bar in the module and specify the .mat file to be read. Unlike “Data File” module, “File Name” module cannot detect the underlying data structure automatically. Manually create a “Group” structure, link the group to the “File Name” module, and one can extract the three variables contained in the file by pressing “Suggest” button. The “3D Path” variable “Var” is also included here.

**Visualize Simulation Data**

DataTank possesses powerful data visualization modules that are compatible with various data formats and real time updates (Fig 1a-1c).

1. To observe numerical data, drag the "3D Path" variable to the "Variable Monitor" panel on the right side of *DataTank*. The positions of all beads at a particular timestep will be displayed in the panel. The simulation time can be toggled in the "Time" sliding bar at the bottom.
2. To create figures, navigate to "Drawing" and create a "3D Space - Rendered" imaging module. A new figure window will be created. Drag the "3D Path" variable into the window, and the chromosome beads and springs will be displayed in the image window. Likewise, one can toggle the "Time" sliding bar to observe chromosome structure at a specific timestep. One can also tweak parameters in the image window to modify visualization features.
3. By default, a "3D Surface" variable has been created in order to represent the nucleus membrane in the figure. It attains the exact parameter settings from the simulation. Drag this variable into the image window to create a sphere around the chromatin chains representing the nucleus membrane. Set the sphere to be semitransparent.

**Analytics Tools**

*Data Analytics Tools for Plasmid Chromosome Simulation*

The file Data_Analysis.tank contains various analytics tools for downstream data analyses applied on outputs from plasmid chromosome simulator. It reads in the raw data following the reading procedure. The data contains not only 3D positional data for every bead, but also customized outputs as time series. For instance, when we focus on the behavior of condensins, the parsed positions for all condensins is included in the output group. Then

appropriate statistical data analytics tools have been developed and progressively under expansion in order to draw further information from the data. Some typical useful modules and the corresponding manual are listed below.

1. Mean Squared Displacement (MSD): Navigate to the variable group "MSD > MSD program", which contains attachment C++ script that computes averaged MSD over a specific portion of the polymer chain. The entire time series was read. The variable "stride" specifies the maximum time step $\tau$ to be analyzed in the program. The parameter "exclude" determines the starting time step for this analysis, since the system is not stable at the beginning stage of the simulation. The output is written in a "1D Plot" format, and can be directly visualized in *DataTank*.
2. Contact Map Analysis: Multiple metrics of contact map was implemented. The approaches can be applied via "Contact map" variable group. As an example, the "Contact frequency" module reads in the bead positions, sample the instances down to a size indicated by variable "sample", and then detects the overall frequency of pairwise distance exceeding the predefined variable "threshold". It outputs the matrix of contact frequency, with the $(i, j)$-entry showing the frequency of close contact between the $i$-th entry and the $j$-th entry.

*Microscope Simulator*

To compare statistical and stochastic properties of simulated data with experimental observations, a *DataTank* script, Microscope_Simulator.tank, convolves the simulation data with a three-dimensional point-spread function to simulate fluorescent microscopy images.

1. Download the script Microscope_Simulator.tank. Compile C++ script in "Microscope Simulator Code" module inside Microscope_Simulator.tank. Add the module as global via clicking the "support" dropdown box and choose "Create as Module…".

Now Microscope Simulator works as a global module and can be executed in all *DataTank* files.

2. To apply the simulator, create a "3D Mesh" variable by navigating to "Variable > 3D > 3D Mesh". Select "Microscope Simulator" in the dropdown box. Three input arguments are required. "Points" takes the list of points to simulate. "Bbox" indicates the bounding box. "Element" reads the mesh basis.

The pipeline outputs the simulated experimental 3D signal from simulation data. One can apply further analysis and compare with experimental observations. Since experimental observations are single image planes in most circumstances, use of the Slicing 3D signal is required.

**Notes**

1. For further help regarding *DataTank*, refer to documentations via "Help". Online help is also available via "Help > Online Help".
2. When compiling attached code through Xcode, Issue may occur failing to compile due to outdated default C++ standard library. To solve this issue, click on the project icon, navigate to "Build Settings > Apple Clang – Language – C++ > C++ Standard Library" and choose "libc++" with C++11 supported. Remote simulation may trigger similar issue because of outdated compiler on server.
3. *ImageTank* is another data visualization tool distributed by Visual Data Tool, Inc., also created by David Adelsteinsson, to further streamline data analysis https://visualdatatools.com/ImageTank/. In addition to all features of *DataTank*, *ImageTank* enhances the user interface, image processing, and computing.

## REFERENCES


Albert, B., Mathon, J., Shukla, A., Saad, H., Normand, C., Léger-Silvestre, I., Villa, D., Kamgoue, A., Mozziconacci, J., Wong, H., Zimmer, C., Bhargava, P., Bancaud, A., & Gadal, O. (2013). Systematic characterization of the conformation and dynamics of budding yeast chromosome XII. *Journal of Cell Biology*, *202*(2), 201–210. https://doi.org/10.1083/jcb.201208186

Bauer, C. R., Hartl, T. A., & Bosco, G. (2012). Condensin II Promotes the Formation of Chromosome Territories by Inducing Axial Compaction of Polyploid Interphase Chromosomes. *PLoS Genet*, *8*(8), 1002873. https://doi.org/10.1371/journal.pgen.1002873

Bloom, K. S. (2008). Beyond the code: The mechanical properties of DNA as they relate to mitosis. In *Chromosoma* (Vol. 117, Issue 2, pp. 103–110). https://doi.org/10.1007/s00412-007-0138-0

Bohn, M., & Heermann, D. W. (2010). Diffusion-driven looping provides a consistent provides a consistent framework for chromatin organization. *PLoS ONE*, *5*(8). https://doi.org/10.1371/journal.pone.0012218

Boisvert, F. M., Van Koningsbruggen, S., Navascués, J., & Lamond, A. I. (2007). The multifunctional nucleolus. In *Nature Reviews Molecular Cell Biology* (Vol. 8, Issue 7, pp. 574–585). https://doi.org/10.1038/nrm2184

Bornfleth, H., Edelmann, P., Zink, D., Cremer, T., & Cremer, C. (1999). Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophysical Journal*, *77*(5), 2871–2886. https://doi.org/10.1016/S0006-3495(99)77119-5

Bowman, G. D., & Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. In *Chemical Reviews* (Vol. 115, Issue 6, pp. 2274–2295). American Chemical Society. https://doi.org/10.1021/cr500350x

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science*, *263*(5148), 802–805. https://doi.org/10.1126/science.8303295

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2001). Capturing Chromosome Conformation. In *science.sciencemag.org*. http://science.sciencemag.org/

Dewar, H., Tanaka, K., Nasmyth, K., & Tanaka, T. U. (2004). Tension between two kinetochores suffices for their bi-orientation on the mitotic spindle. *Nature*, *428*(6978), 93–97. https://doi.org/10.1038/nature02328

Dion, V., & Gasser, S. M. (2013). Chromatin movement in the maintenance of genome stability. In *Cell* (Vol. 152, Issue 6, pp. 1355–1364). https://doi.org/10.1016/j.cell.2013.02.010

Eriksson, P. R., Mendiratta, G., Mclaughlin, N. B., Wolfsberg, T. G., Mariño-Ramírez, L., Pompa, T. A., Jainerin, M., Landsman, D., Shen, C.-H., & Clark, D. J. (2005). Global Regulation by the Yeast Spt10 Protein Is Mediated through Chromatin Structure and the

Histone Upstream Activating Sequence Elements. *MOLECULAR AND CELLULAR BIOLOGY*, *25*(20), 9127–9137. https://doi.org/10.1128/MCB.25.20.9127-9137.2005

Faller, R., & Müller-Plathe, F. (2001). Chain stiffness intensifies the reptation characteristics of polymer dynamics in the melt. *ChemPhysChem*, *2*(3), 180–184. https://doi.org/10.1002/1439-7641(20010316)2:3<180::AID-CPHC180>3.0.CO;2-Z

Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., & Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science*, *360*(6384), 102–105. https://doi.org/10.1126/science.aar7831

Hancock, R. (2012). Structure of metaphase chromosomes: A role for effects of macromolecular crowding. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0036045

Handwerger, K. E., Murphy, C., & Gall, J. G. (2003). Steady-state dynamics of Cajal body components in the Xenopus germinal vesicle. *Journal of Cell Biology*, *160*(4), 495–504. https://doi.org/10.1083/jcb.200212024

Hirano, T., Kobayashi, R., & Hirano, M. (1997). Condensins, chromosome condensation protein complexes containing XCAP- C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. *Cell*, *89*(4), 511–521. https://doi.org/10.1016/S0092-8674(00)80233-0

Hult, C., Adalsteinsson, D., Vasquez, P. A., Lawrimore, J., Bennett, M., York, A., Cook, D., Yeh, E., Forest, M. G., & Bloom, K. (2017). Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus. *Nucleic Acids Research*, *45*(19), 11159–11173. https://doi.org/10.1093/nar/gkx741

Johansen, K. M., & Johansen, J. (2006). Regulation of chromatin structure by histone H3S10 phosphorylation. *Springer*, *14*(4), 393–404. https://doi.org/10.1007/s10577-006-1063-4

Kim, E., Kerssemakers, J., Shaltiel, I. A., Haering, C. H., & Dekker, C. (2020). DNA-loop extruding condensin complexes can traverse one another. *Nature*, *579*(7799), 438–442. https://doi.org/10.1038/s41586-020-2067-5

Lawrimore, J, Aicher, J., Hahn, P., … A. F.-M. biology of, & 2016, undefined. (2016). ChromoShake: a chromosome dynamics simulator reveals that chromatin loops stiffen centromeric chromatin. *Am Soc Cell Biol*, *27*(1), 153–166. https://doi.org/10.1091/mbc.E15-08-0575

Lawrimore, Josh, Friedman, B., Doshi, A., & Bloom, K. (2017). RotoStep: A Chromosome Dynamics Simulator Reveals Mechanisms of Loop Extrusion. *Cold Spring Harbor Symposia on Quantitative Biology*, *1*. https://doi.org/10.1101/sqb.2017.82.033696

Lawrimore, Josh, He, Y., Forest, G. M., & Bloom, K. (2019). Three-dimensional thermodynamic simulation of condensin as a DNA-based translocase. In *Methods in Molecular Biology* (Vol. 2004, pp. 291–318). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9520-2_21

Li, Z., Vizeacoumar, F. J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F. S., Min, R., Vandersluis, B., Bellay, J., Devit, M., Fleming, J. A., Stephens, A., Haase, J., Lin, Z.-Y.,

Baryshnikova, A., Lu, H., Yan, Z., Jin, K., Barker, S., … Boone, C. (2011). systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nature Biotechnology*, *29*. https://doi.org/10.1038/nbt.1832

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luo, J., Deng, X., Buehl, C., Xu, X., & Kuo, M.-H. (2016). Identification of Tension Sensing Motif of Histone H3 in Saccharomyces cerevisiae and Its Regulation by Histone Modifying Enzymes. *Genetics*, *204*, 1029–1043. https://doi.org/10.1534/genetics.116.192443

Marko, J F, & Siggia, E. D. (1995). Statistical mechanics of supercoiled DNA. In *PHYSICAL REVIEW E VOLUME* (Vol. 52). https://journals.aps.org/pre/abstract/10.1103/PhysRevE.52.2912

Marko, John F., & Siggia, E. D. (1994). Fluctuations and supercoiling of DNA. *Science*, *265*(5171), 506–508. https://doi.org/10.1126/science.8036491

Mikus, M. D., & Petes, T. D. (1982). Recombination between genes located on nonhomologous chromosomes in Saccharomyces cerevisiae. *Genetics*, *101*(3–4), 369–404. https://www.genetics.org/content/101/3-4/369.short

Morris, T. D., Weber, L. A., Hickey, E., Stein, G. S., & Stein2, J. L. (1991). Changes in the Stability of a Human H3 Histone mRNA during the HeLa Cell Cycle. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 11, Issue 1). http://mcb.asm.org/

Natsoulis, G., Winston, F., & Boeke, J. D. (1994). The SPT10 and SPT21 genes of Saccharomyces cerevisiae. *Genetics*, *136*(1), 93–105. https://www.genetics.org/content/136/1/93.short

Neice, A. (2010). Methods and Limitations of Subwavelength Imaging. In *Advances in Imaging and Electron Physics* (Vol. 163, Issue C, pp. 117–140). Academic Press Inc. https://doi.org/10.1016/S1076-5670(10)63003-0

Oakes, M. L., Johzuka, K., Vu, L., Eliason, K., & Nomura, M. (2006). Expression of rRNA Genes and Nucleolus Formation at Ectopic Chromosomal Sites in the Yeast Saccharomyces cerevisiae ‡. *MOLECULAR AND CELLULAR BIOLOGY*, *26*(16), 6223–6238. https://doi.org/10.1128/MCB.02324-05

Oshidari, R., Strecker, J., Chung, D. K. C., Abraham, K. J., Chan, J. N. Y., Damaren, C. J., & Mekhail, K. (2018). Nuclear microtubule filaments mediate non-linear directional motion of chromatin and promote DNA repair. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05009-7

Peskin, C. S. (2002). The immersed boundary method. *Acta Numerica*, *11*, 479–517.

Quammen, C. W., Richardson, A. C., Haase, J., Harrison, B. D., Taylor, R. M., II, & Bloom, K. S. (2008). FluoroSim: A Visual Problem-Solving Environment for Fluorescence Microscopy. *Eurographics Workshop on Visual Computing for Biomedicine*, *2008*, 151. https://doi.org/10.2312/VCBM/VCBM08/151-158

Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., & Belmont, A. S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. In *Journal of Cell Biology* (Vol. 135, Issue 6 II). https://doi.org/10.1083/jcb.135.6.1685

Rubinstein, M., & Colby, R. H. (2003). *Polymer physics*. Oxford university press New York.

Sakai, Y., Mochizuki, A., Kinoshita, K., Hirano, T., & Tachikawa, M. (2018). Modeling the functions of condensin in chromosome shaping and segregation. *PLoS Computational Biology*, *14*(6). https://doi.org/10.1371/journal.pcbi.1006152

Sati, S., & Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. In *Chromosoma* (Vol. 126, Issue 1, pp. 33–44). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00412-016-0593-6

Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M. L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., … Campbell, P. J. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, *144*(1), 27–40. https://doi.org/10.1016/j.cell.2010.11.055

Straight, A. F., Belmont, A. S., Robinett, C. C., & Murray, A. W. (1996). GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Current Biology*, *6*(12), 1599–1608. https://doi.org/10.1016/S0960-9822(02)70783-5

Supek, F., & Lehner, B. (2017). Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell*, *170*(3), 534-547.e23. https://doi.org/10.1016/j.cell.2017.07.003

Teif, V. B., & Clarkson, C. T. (2019). Nucleosome positioning. *Encyclopedia of Bioinformatics and Computational Biology*, *2*, 308–317.

Terakawa, T., Bisht, S., Eeftens, J. M., Dekker, C., Haering, C. H., & Greene, E. C. (2017). The condensin complex is a mechanochemical motor that translocates along DNA. *Science*, *358*(6363), 672–676. https://doi.org/10.1126/science.aan6516

Thadani, R., Uhlmann, F., & Heeger, S. (2012). Condensin, chromatin crossbarring and chromosome condensation. In *Current Biology* (Vol. 22, Issue 23, pp. R1012–R1021). Cell Press. https://doi.org/10.1016/j.cub.2012.10.023

Therizolsa, P., Duong, T., Dujon, B., Zimmer, C., & Fabre, E. (2010). Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(5), 2025–2030. https://doi.org/10.1073/pnas.0914187107

Vasquez, P. A., Hult, C., Adalsteinsson, D., Lawrimore, J., Forest, M. G., & Bloom, K. (2016). Entropy gives rise to topologically associating domains. *Nucleic Acids Research*, *44*(12), 5540–5549. https://doi.org/10.1093/nar/gkw510

Walker, B., Taylor, D., Lawrimore, J., Hult, C., Adalsteinsson, D., Bloom, K., & Gregory

Forest, M. (2019). Transient crosslinking kinetics optimize gene cluster interactions. *PLoS Computational Biology*, *15*(8). https://doi.org/10.1371/journal.pcbi.1007124

Weber, S. C., Spakowitz, A. J., & Theriot, J. A. (2012). Nonthermal ATP-dependent fluctuations contribute to the in vivo motion of chromosomal loci. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7338–7343. https://doi.org/10.1073/pnas.1119505109

Wong, H., Arbona, J. M., & Zimmer, C. (2013). How to build a yeast nucleus. *Nucleus (United States)*, *4*(5). https://doi.org/10.4161/nucl.26226

Yang, X., Shen, X., Long, J., & Chen, H. (2012). An Improved Median-based Otsu Image Thresholding Algorithm. *AASRI Procedia*, *3*, 468–473. https://doi.org/10.1016/j.aasri.2012.11.074

Young, W. C., Budynas, R. G., York, N., San, C., Lisbon, F., Madrid, L., City, M., New, M., San, D., & Seoul, J. (1976). *Roark's Formulas for Stress and Strain*. http://198.24.197.162/reference/Roark Young.pdf

Zimmer, C. (2018). Chromatin mobility upon DNA damage: state of the art and remaining questions. *Springer*, *65*(1). https://doi.org/10.1007/s00294-018-0852-6